

Jelölésjegyzék

Az alábbi táblázat tartalmazza a könyvben használt fontosabb jelöléseket. Amennyiben ettől eltérünk, azt külön jelezzük.

\mathbb{R}	a valós számok teste
\mathbb{N}	természetes számok halmaza
$\mathbf{A} \in \mathbb{R}^{N \times M}, a_{ij}$	$N \times M$ méretű valós mátrix, ill. i -edik sorának j -edik eleme
$\mathbf{v} = \langle v_1, \dots, v_n \rangle \in \mathbb{R}^n$	n elemű (valós) vektor
$\mathbf{v} \in \mathbb{R}^{1 \times n}, \mathbf{w} \in \mathbb{R}^{n \times 1}$	sorvektor, illetve oszlopvektor (ha hangsúlyozni akarjuk a vektor alakját)
$\langle \mathbf{u}, \mathbf{v} \rangle$	\mathbf{u} és \mathbf{v} vektorok skalárszorzata
$ A $	A halmaz elemszáma
$c; c_j \in C$	kategória; a kategóriarendszer egy eleme
\mathbf{c}	a c kategória kategóriaprofilját megadó vektor
$C = \{c_1, \dots, c_{ C }\}$	kategóriák halmaza
cf_k	a t_k szó gyűjteménytámogatottsága
$d; d_i \in D$	dokumentum; a dokumentumgyűjtemény egy eleme
\mathbf{d}	a d dokumentum vektorrepresentációja
$D = \{d_1, \dots, d_N\}$	dokumentumgyűjtemény (korpusz) és elemei
$d(\cdot, \cdot)$	távolságfüggvény
df_k	a t_k szó dokumentumgyakorisága a korpuszban
l_d, l_t	tanító-, ill. tesztdokumentumok átlagos vektormérete (ritka vektorként)
L_d, L_t	tanító-, ill. tesztdokumentumok átlagos hossza (szavak száma)
M, N	egyedi szavak, ill. dokumentumok száma
n_k	a korpusz t_k szót tartalmazó dokumentumainak száma
n_{ki}	a t_k szó előfordulásainak száma d_i dokumentumban
Neg_j	a c_j kategóriába nem tartozó tanítóadatok
Pos_j	a c_j kategóriába tartozó tanítóadatok
$s(\cdot, \cdot)$	hasonlóságfüggvény
$S_k(\mathbf{d}_i)$	\mathbf{d}_i -hez legközelebbi k szomszéd halmaza
t, t_k	szó (terminus); a vektortér k -edik dimenziójához

Fontosabb szakkifejezések rövidítésekkel magyarul és angolul

magyar	angol	rövidítés
adaptív szűrés	adaptive filtering	
alulról-felfelé	bottom-up	
alultövezési index	under-stemming index	UI
anaforafeloldás	anaphora resolution	AR
aratórobot	harvester	
átlagos kapcsolódás	group-average link	
balelemző	top-down parser	
csomópont	node	
dokumentumszűrés	text filtering	
dokumentumvezérelt osztályozás	document-pivoted categorization	DPC
döntési fa alapú osztályozó	decision tree classifier	DT-classifier
döntési szabály alapú osztályozó	decision rule classifier	DR-classifier
dzsókerkarakter	wildcard	
egycímkés osztályozás	single-label classification	
egyszerű kapcsolódás	single-link	
eltolás	bias	
erőforrás-leíró keretrendszer	resource description framework	RDF
feldolgozási folyamat	stream	
feltételes valószínűségi mező	conditional random fields	CRF
felügyelet nélküli tanulás	unsupervised learning	
felügyelt tanulás	supervised learning	
felülről-lefelé	top-down	
fokozatos tanulás	incremental learning	
fontossági forrás	source of rank	
főkomponens-analízis	principal components analysis	PCA
frázissablon	phrasal template	
gyűjteménytámogatottság	collection frequency	CF
hibavezérelt tanulás	mistake driven learning	
hierarchikus (szöveg)osztályozás	hierarchical text categorization	HTC
információkinyerés	information extraction	IE
információnyereség	information gain	IG
információ-visszakeresés	information retrieval	IR
jellemzőkinyerés	term extraction	
jellemzőkiválasztás	term selection	
jobbelemző	bottom-up parser	
k -átlag	k -means	
kategóriavezérelt osztályozás	category-pivoted categorization	CPC
kereszthivatkozás	co-reference	
kereszthivatkozás-feloldás	co-reference resolution	

magyar	angol	rövidítés
keresztvalidáció, k -szoros keret	cross-validation, k -fold frame	
kéretlen levelek szűrése	spam filtering	
kifejezéssablon	phrasal template	
kiterjesztett vagy bővített átmenet-háló	augmented transition network	ATN
kiválasztási elv	selection policy	
kötegelt tanulás	batch learning	
látens szemantikus indexelés	latent semantic indexing	LSI
legközelebbi szomszéd osztályozó (k -NN osztályozó)	nearest neighbor classifier	k -NN
lineáris legkisebb négyzetek módszer	linear least-squares fit	LLSF
lusta tanuló	lazy learner	
maximum entrópia Markov-modell	maximum entropy Markov modell	MEMM
meredekségi faktor	slope factor	
metszés (döntési fáé)	pruning	
minta alapú osztályozó	example-based classifier	
mintaillesztés	pattern matching	
névelem-felismerés	named entity recognition	NER
nyelő	rank sink	
nyelvközi információkinyerés	cross-language information extraction	CLIE
osztályozó bizottság	classifier committee, ensemble classifier	
öregedési algoritmus	aging algorithm	
összegzőkészítő eljárás	text summarization method	
párhuzamos feldolgozási elv	parallelization policy	
pillanatkép	snapshot	
radiális bázisfüggvény	radial basis function	RBF
rangsoroló eljárás	ranking algorithm	
rejtett Markov-modell	hidden Markov model	HMM
relevancia-visszacsatolás	relevance feedback	
szekvencia alapú modell	structured prediction	SP
szinguláris értékfelbontás	singular value decomposition	SVD
szó-dokumentum mátrix	term-document matrix	TD matrix
szógyakoriság alapú súlyozás (TF-súlyozás)	term frequency	TF
szótövező	stemmer	
szózsákmodell	bag of words model	

magyar	angol	rövidítés
szövegosztályozás	text categorization/classification	TC
szupportvektor gép	support vektor machine	SVM
támogató osztályozás	categorization assistance	
tanítóhalmaz	training set	
teljes kapcsolódás	complete-link	
természetes nyelvek megértése	natural language understanding	NLU
természetes nyelvű adatbázis-interfész	natural language interfaces to data-bases	NLIDB
természetes nyelvű mélyhálókere-ső-interfész	natural language interface to deep web searcher	NLIDW
terminusfrekvencia és inverz dokumentumfrekvencia	term frequency & inverse document frequency	tf-idf
teszthalmaz	test set	
tisztaság	purity	
többcímkes osztályozás	multi-label classification	
többértelmű szavak egyértelműsítése	word sense disambiguation	
többségi döntés	majority voting	
többszintes osztályozás	multi-level classification	
válaszkereső rendszerek	question answering systems	QAS
udvariassági elv	politeness policy	
ugró pointer	skip pointer	
újralátogatási elv	re-visit policy	
újraparametrizálás	re-parametrization	
úrlap/nyomtatvány	form	

Egyéb alkalmazott angol rövidítések, és az esetlegesen kapcsolódó honlapcímek

rövidítés	jelentés	URL
ACE	Automatic Content Extraction	www.itl.nist.gov/iad/894.01/tests/ace/
ANSI	American National Standards Institute	www.ansi.org
CART	Classification and Regression Trees	www.salfordsystems.com/cart.php
CoNLL	Conference on Computational Natural Language Learning	ifarm.nl/signll/conll/
ETO	Egyetemes Tizedes Osztályozás	
HITEC	Hierarchical TEXT Categorizer	categorizer.tmit.bme.hu
ID3	Interactive Dichotomizer 3	
IPC	International Patent Classification (Nemzetközi Szabadalmi Osztályozás)	www.wipo.int/classifications/ipc/en/
ISO	International Organization for Standardization	www.iso.org
KWIC	Key Word in Context	
MUC	Message Understanding Conferences	www.itl.nist.gov/iaui/894.02/related_projects/muc/
OPAC	Open Public Access Catalog	
SMART	Salton's Magical Automatic Retriever of Text	
SQL	Structured Query Language	www.ncb.ernet.in/education/modules/dbms/sql99index.html
TREC	Text REtrieval Conference	trec.nist.gov
WIPO	World Intellectual Property Organization (Nemzetközi Szellemi Tulajdonok Szervezet)	www.wipo.int