SIEMENS

Data- and model-driven approaches to finding semantic representations in fMRI data

Francisco Pereira

Imaging and Visualization

Siemens Corporation,

Corporate Research & Technology

Matthew Botvinick, Greg Detre

Psychology Department

Princeton University

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

current areas of work

SIEMENS

- neuroimaging (Mariappan Nadar's group)
  - functional connectivity, resting state analysis, tissue classification
- interventional imaging (Christophe Chefd'hotel's group)
  - image segmentation, tissue classification

- imaging analysis methods (Matthew Botvinick and Uri Hasson)
  - searchlight classification toolbox
  - searchlight similarity structure toolbox
  - multi-condition functional connectivity
- representation of semantic information (Matthew Botvinick)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

outline

SIEMENS

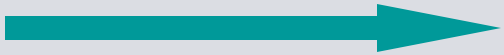
- machine learning classifiers in fMRI
- a method for systematically extracting information across the brain for use in a classifier
- using text data and matrix factorization to find representations of semantic information in fMRI

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

two different questions

SIEMENS

GLM

stimulus  fMRI activation

(single voxel)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

fMRI analysis with a classifier

SIEMENS

[Kamitani&Tong, 2005]

subjects see gratings in one of 8 orientations

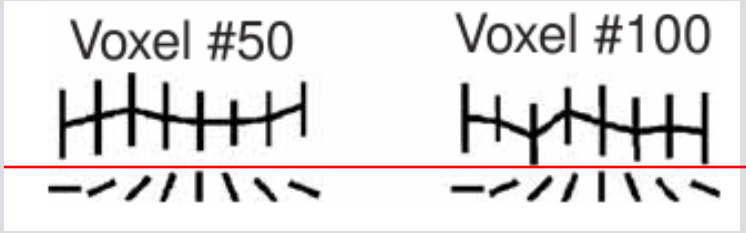
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

fMRI analysis with a classifier

SIEMENS

[Kamitani&Tong, 2005]

subjects see gratings in one of 8 orientations



voxels in visual cortex respond similarly to different orientations

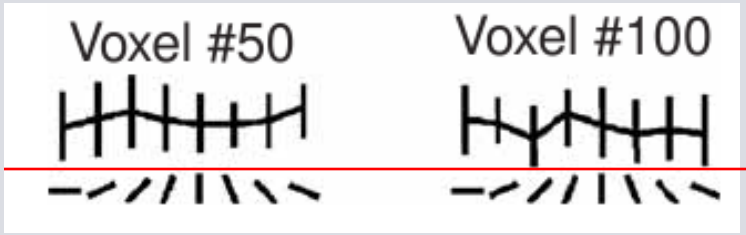
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

fMRI analysis with a classifier

SIEMENS

[Kamitani&Tong, 2005]

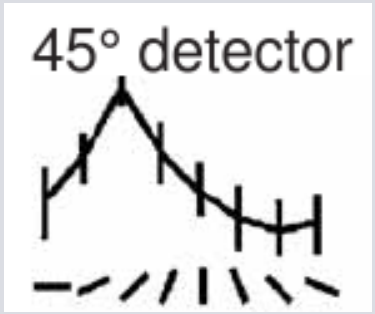
subjects see gratings in one of 8 orientations



voxels in visual cortex respond similarly to different orientations

classifier

yet, voxels can be combined to predict the orientation of the grating being seen!



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

two different questions

SIEMENS

GLM

stimulus → fMRI activation

(single voxel)

Classifier

stimulus ← fMRI activation

(multiple voxels)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classifiers on fMRI, mind-reading, etc

SIEMENS

We can predict!

[Mitchell et al 2004, Haynes/Rees 2006, Norman et al 2006, Pereira et al 2009,...]

- what is the orientation of a stimulus visual grating?
- is the subject seeing a sentence or a picture?
- which of several categories of words or pictures is a subject seeing?
- is the subject reading an ambiguous sentence?
- same or different sentence?
- what is the subject perceiving?
- is the subject concealing information?
- “do you know this terrorist, sir?”
- “is there still a fire for me in his brain?”

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classifiers on fMRI, mind-reading, etc

SIEMENS

We can predict!

[Mitchell et al 2004, Haynes/Rees 2006, Norman et al 2006, Pereira et al 2009,...]

- what is the orientation of a stimulus visual grating?
- is the subject seeing a sentence or a picture?
- which of several categories of words or pictures is a subject seeing?
- is the subject reading an ambiguous sentence?
- same or different sentence?
- what is the subject perceiving?
- is the subject concealing information?
- “do you know this terrorist, sir?”
- “is there still a fire for me in his brain?”

creepiness gradient

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

two different questions revisited

SIEMENS

GLM

stimulus → fMRI activation

(single voxel)

Classifier

stimulus ← fMRI activation

(multiple voxels)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

two different questions revisited

SIEMENS

GLM

stimulus → fMRI activation
(single voxel)

Classifier

stimulus ← fMRI activation
(multiple voxels)

stimulus → mind → brain → fMRI activation
(multiple voxels)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

two different questions revisited

SIEMENS

GLM

stimulus → fMRI activation

(single voxel)

Classifier

stimulus ← fMRI activation

(multiple voxels)

stimulus → mind → brain → fMRI activation

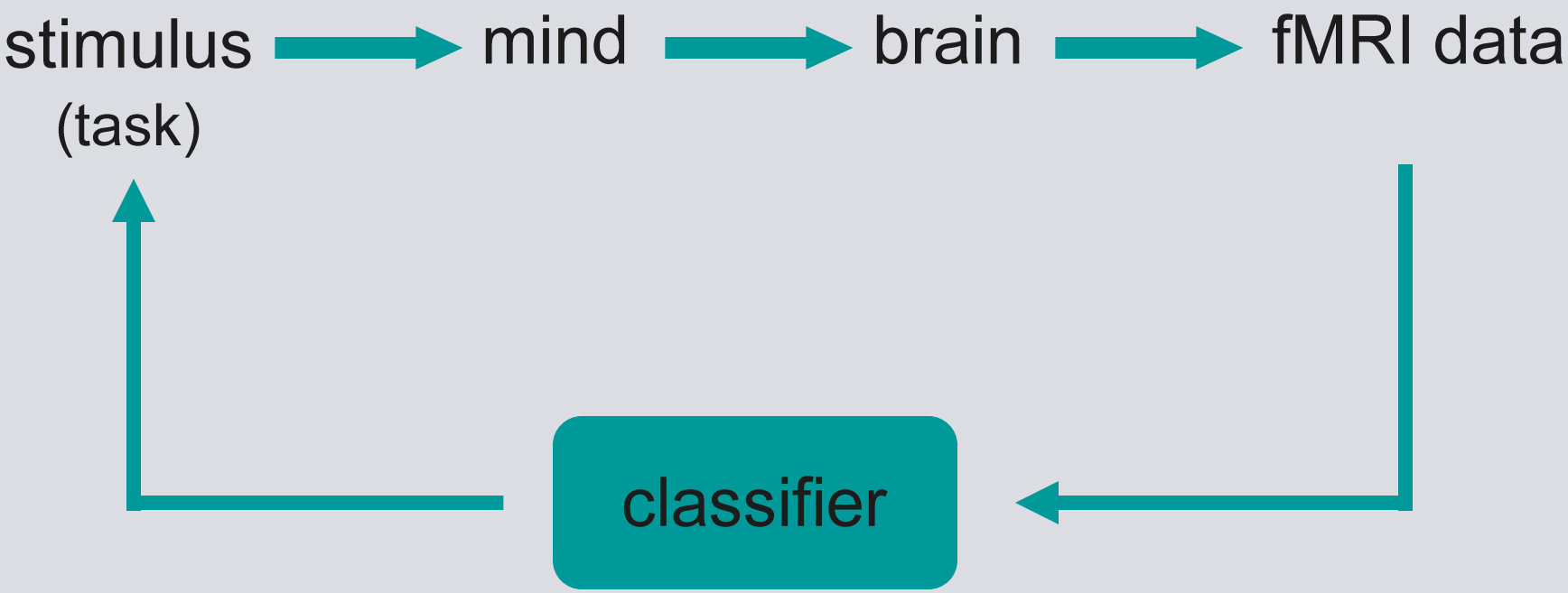
(multiple voxels)

(not an endorsement of dualism)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where are we?

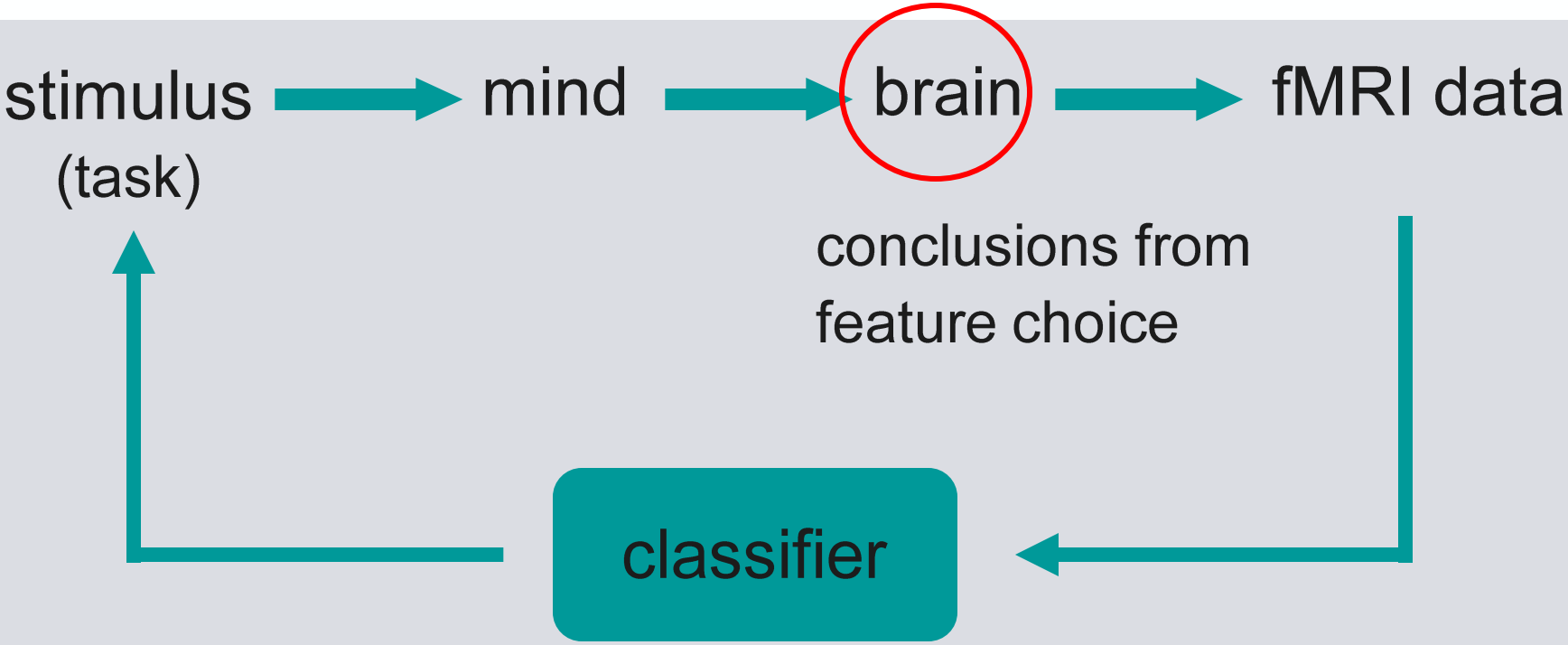
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where are we?

SIEMENS



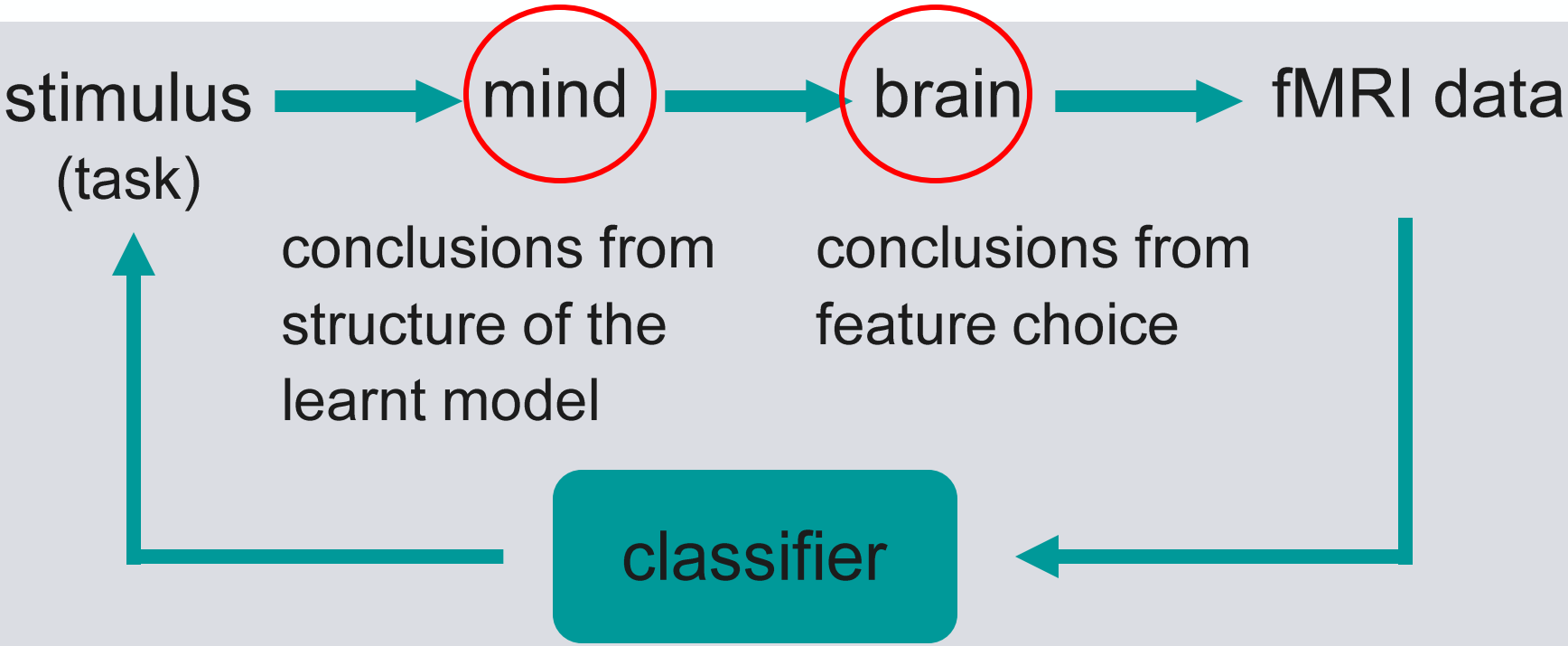
- voxel location
- voxel behaviour
- time within trial

dependent on experiment

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where are we?

SIEMENS



- weights on features
- hidden layer activations

dependent on prediction model

- voxel location
- voxel behaviour
- time within trial

dependent on experiment

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

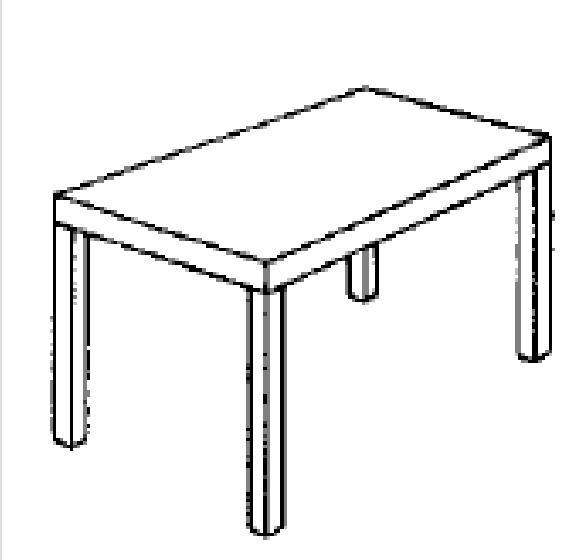
12 categories experiment

SIEMENS

task:

- stimulus: word + drawing
- subject visualizes object, thinks of properties, using it, etc
- 3 seconds per trial

[Mitchell et al, 2008]



Table

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

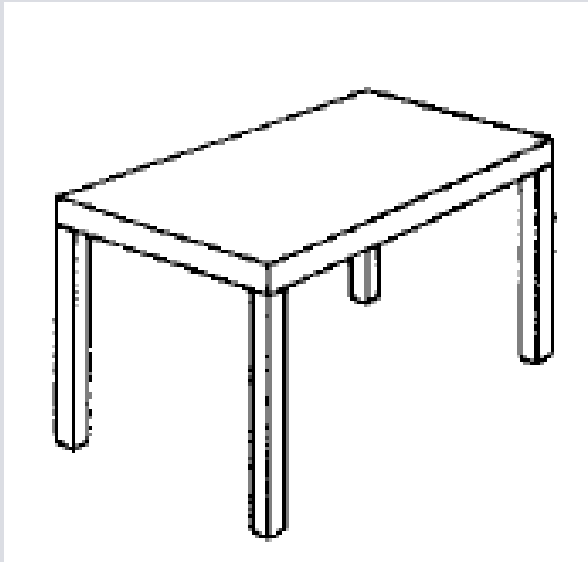
12 categories experiment

SIEMENS

[Mitchell et al, 2008]

task:

- stimulus: word + drawing
- subject visualizes object, thinks of properties, using it, etc
- 3 seconds per trial



Table

dataset:

- 12 categories, 5 exemplars of each
- 6 epochs with all 60 exemplars
- 360 examples (average image at trial peak)
- even/odd epoch cross-validation (and leave-one-out inside)

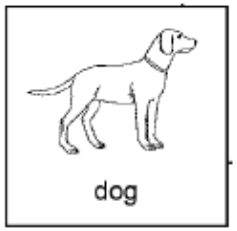
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

12 categories experiment

SIEMENS

[Mitchell et al, 2008]

60 exemplars



Categories	Exemplars				
BODY PARTS	leg	arm	eye	foot	hand
FURNITURE	chair	table	bed	desk	dresser
VEHICLES	car	airplane	train	truck	bicycle
ANIMALS	horse	dog	bear	cow	cat
KITCHEN UTENSILS	glass	knife	bottle	cup	spoon
TOOLS	chisel	hammer	screwdriver	pliers	saw
BUILDINGS	apartment	barn	house	church	igloo
PART OF A BUILDING	window	door	chimney	closet	arch
CLOTHING	coat	dress	shirt	skirt	pants
INSECTS	fly	ant	bee	butterfly	beetle
VEGETABLES	lettuce	tomato	carrot	corn	celery
MAN MADE OBJECTS	refrigerator	key	telephone	watch	bell

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classifier examples

SIEMENS

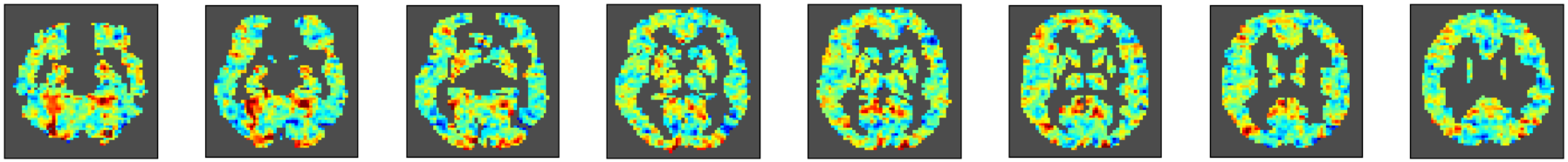
- 2 category problem: tools vs buildings

- the usual approaches
  - whole-brain classifiers
  - feature selection
  - searchlight classifiers

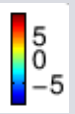
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

the usual approaches 1

SIEMENS



classifier weights



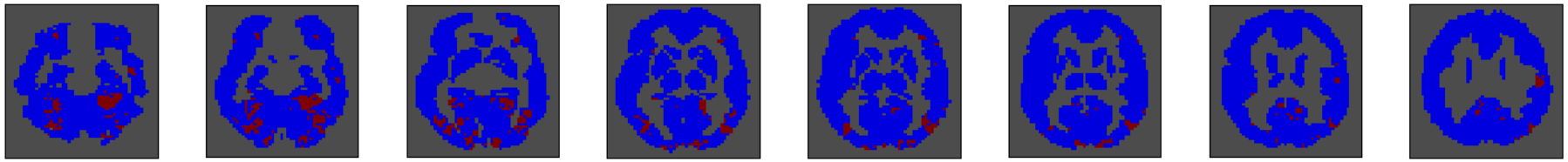
whole-brain classifiers (tools vs buildings)

- plot the weight of each voxel in a linear classifier
- noisy
- poor performance in multi-class situations

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

the usual approaches 2

SIEMENS



selected features

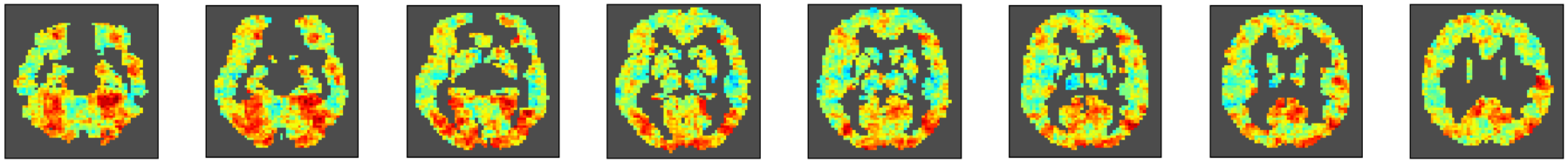
feature selection (tools vs buildings)

- build a classifier using only selected voxels, plot the ones chosen
- poor reproducibility over cross-validation folds
- selects only some features (up to #, method, etc)

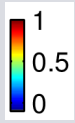
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

the usual approaches 3

SIEMENS



classifier accuracy



searchlight classifiers (tools vs buildings)

- train one classifier per 3x3x3 voxel patch
- plot accuracy of that at the patch center voxel
- localized information
- limited number of possible class distinctions in a single one

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

outline

SIEMENS

- machine learning classifiers in fMRI
- a method for systematically extracting information across the brain for use in a classifier
- using text data and matrix factorization to find representations of semantic information in fMRI

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?

categories can be distinguished

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?
categories can be distinguished

- a category from all others

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?
categories can be distinguished

- a category from all others
  - GLM-ish idea
  - not reasonable with many categories (or exemplars, or ...)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?
categories can be distinguished

- a category from all others
  - GLM-ish idea
  - not reasonable with many categories (or exemplars, or ...)
- a category from another category

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

what does the presence of information mean?

categories can be distinguished

- a category from all others
  - GLM-ish idea
  - not reasonable with many categories (or exemplars, or ...)
- a category from another category
  - hypothesis: at fMRI resolution, there are always small regions

    where the pattern elicited by two categories is different

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

re-thinking the whole thing

SIEMENS

[Pereira&Botvinick 2011, in preparation]

a method for

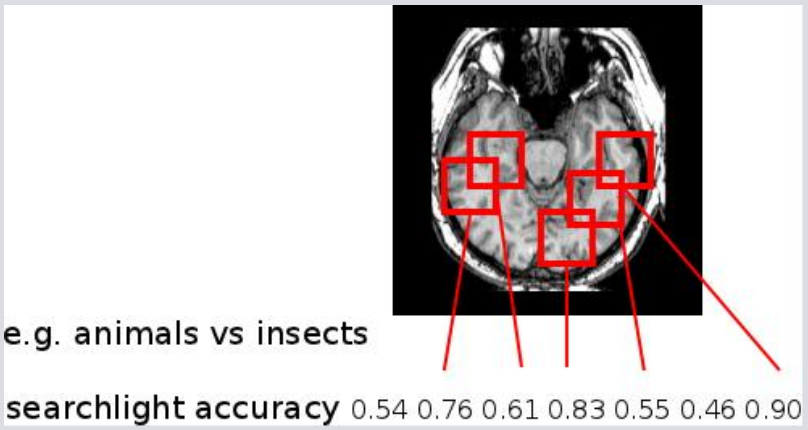
- finding as much information as possible
- combining it to build a better classifier
- using it to produce information maps

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

find informative searchlights

SIEMENS

consider all searchlights
and cross-validate a classifier
for each pair of categories

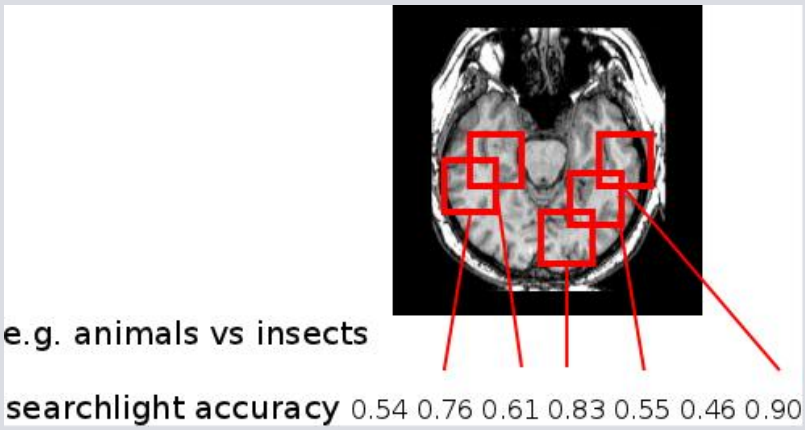


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

find informative searchlights

SIEMENS

consider all searchlights
and cross-validate a classifier
for each pair of categories



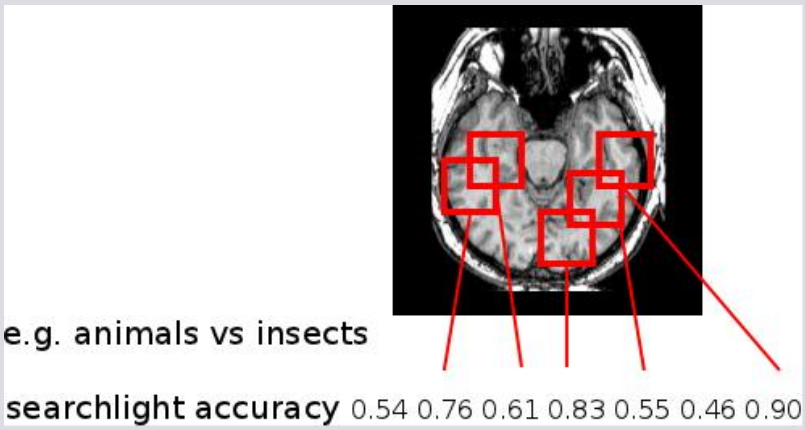
identify the significant results
(false discovery rate 0.01)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

find informative searchlights

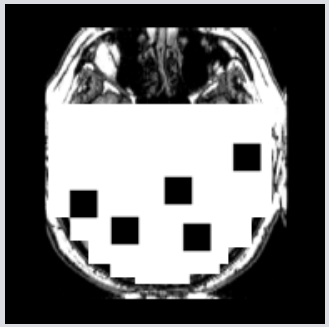
SIEMENS

consider all searchlights
and cross-validate a classifier
for each pair of categories



identify the significant results
(false discovery rate 0.01)

arrange them into a
significance image

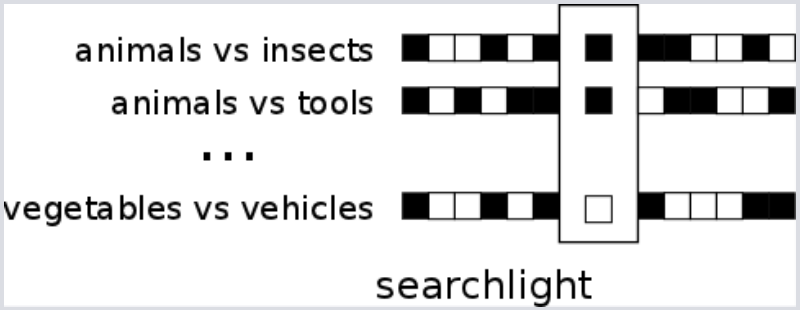


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

aggregate information

SIEMENS

this is done for all 66
pairwise classification tasks

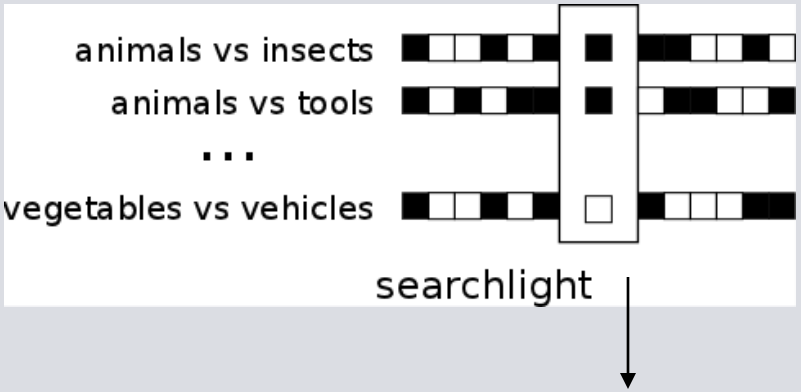


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

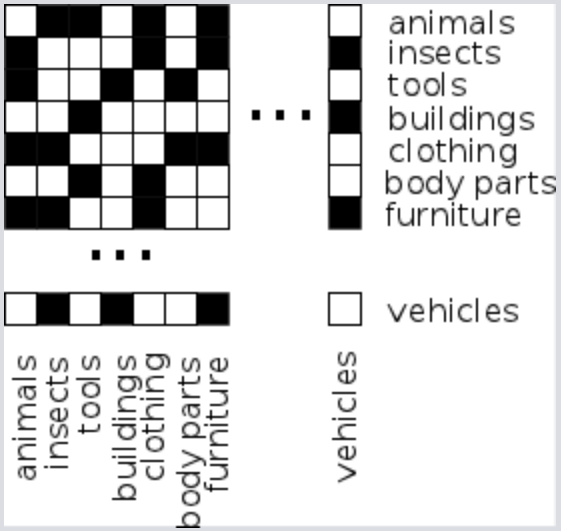
aggregate information

SIEMENS

this is done for all 66 pairwise classification tasks



and the results for each searchlight become a binary “confusion matrix” of supported distinctions

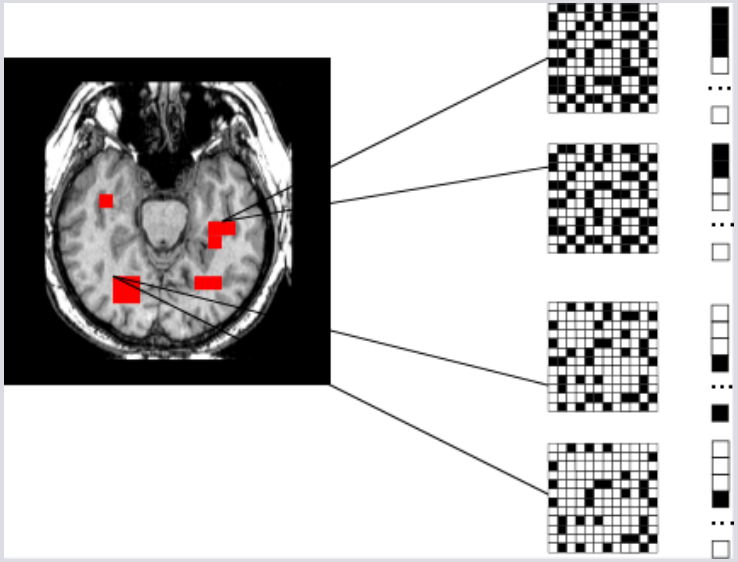


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

SIEMENS

the confusion matrices
can be clustered spatially
into groups supporting
similar category distinctions

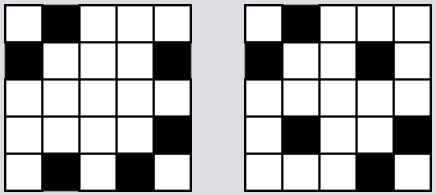


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

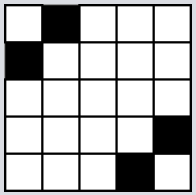
SIEMENS

for two matrices consider



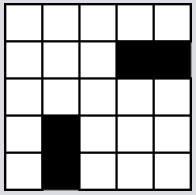
$\mathbf{v}_i$ $\mathbf{v}_j$

matches



$\sum_{pairs} \mathtt{AND}(\mathbf{v}_i, \mathbf{v}_j)$

mismatches



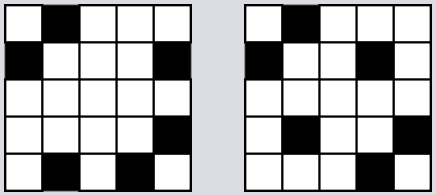
$\sum_{pairs} \mathtt{XOR}(\mathbf{v}_i, \mathbf{v}_j)$

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

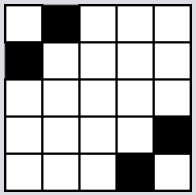
SIEMENS

for two matrices consider



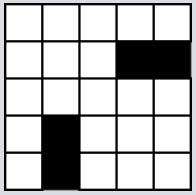
$\mathbf{v}_i$ $\mathbf{v}_j$

matches



$\sum_{pairs} \mathtt{AND}(\mathbf{v}_i, \mathbf{v}_j)$

mismatches



$\sum_{pairs} \mathtt{XOR}(\mathbf{v}_i, \mathbf{v}_j)$

similarity measure

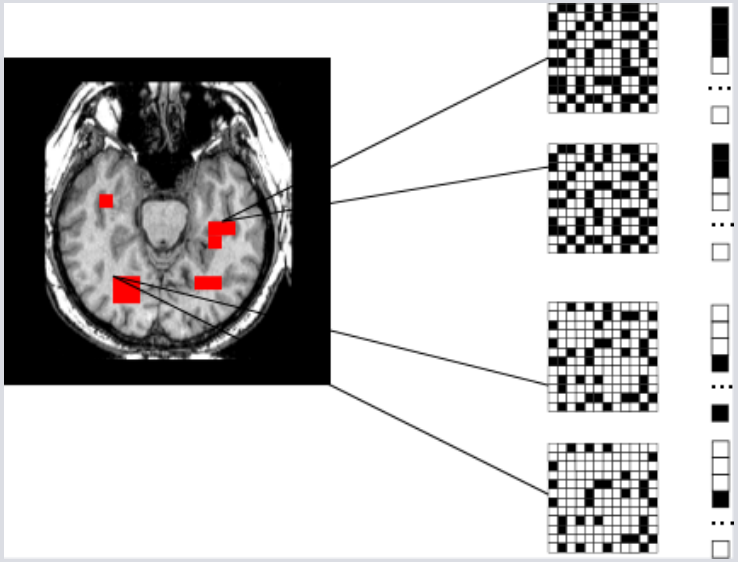
$$\text{similarity}(\mathbf{v}_i, \mathbf{v}_j) = \frac{\sum_{pairs} \mathtt{AND}(\mathbf{v}_i, \mathbf{v}_j) - \frac{\sum_{pairs} \mathtt{XOR}(\mathbf{v}_i, \mathbf{v}_j)}{2}}{\sum_{pairs} \mathtt{AND}(\mathbf{v}_i, \mathbf{v}_j)}$$

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

SIEMENS

the confusion matrices
can be clustered spatially
into groups supporting
similar category distinctions

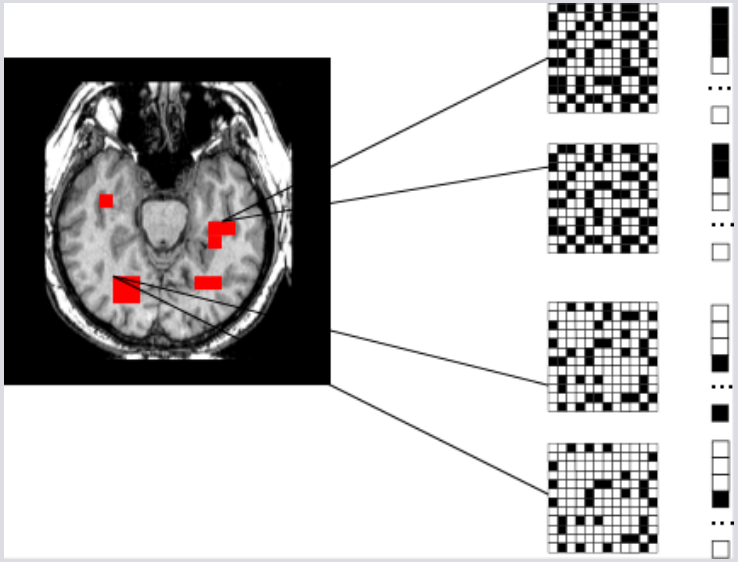


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

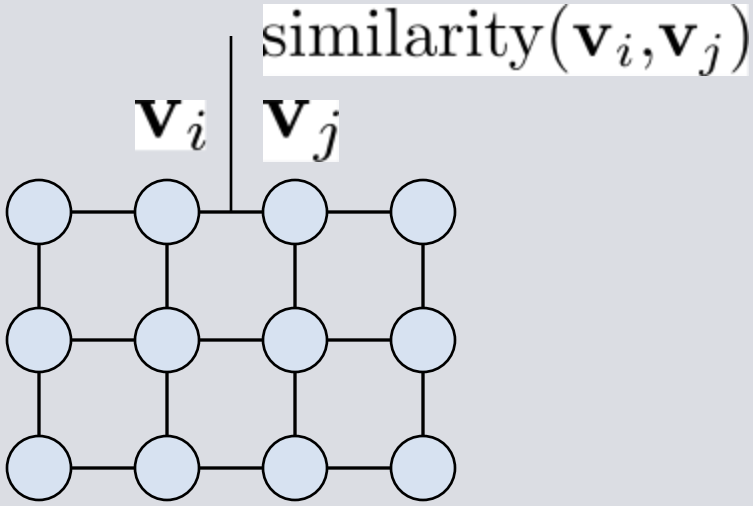
cluster searchlights into patches

SIEMENS

the confusion matrices
can be clustered spatially
into groups supporting
similar category distinctions



representing each ROI as
a weighted similarity graph

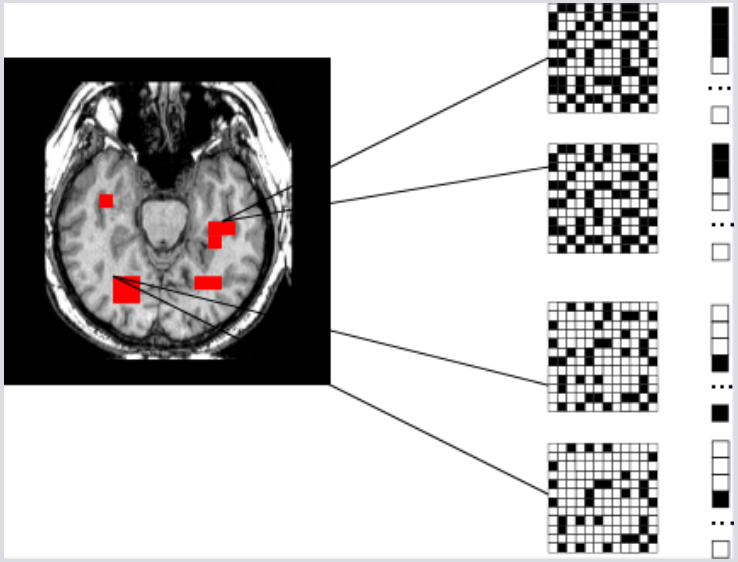


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

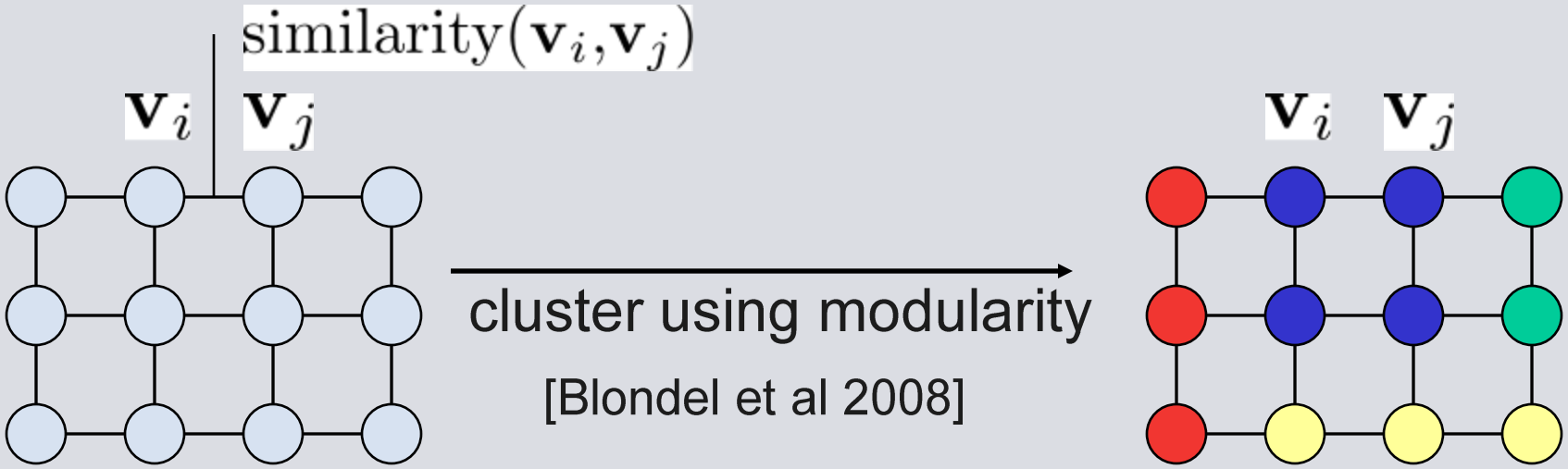
cluster searchlights into patches

SIEMENS

the confusion matrices
can be clustered spatially
into groups supporting
similar category distinctions



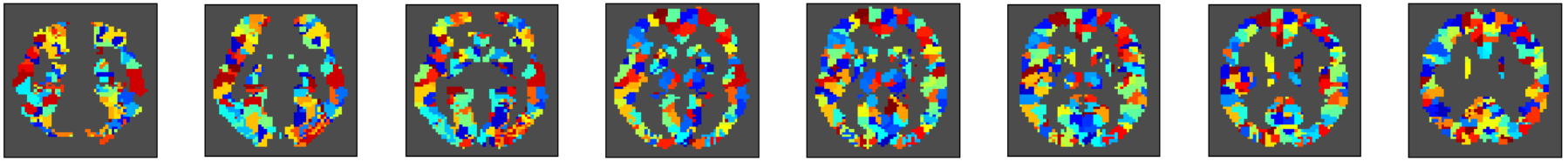
representing each ROI as
a weighted similarity graph



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

SIEMENS



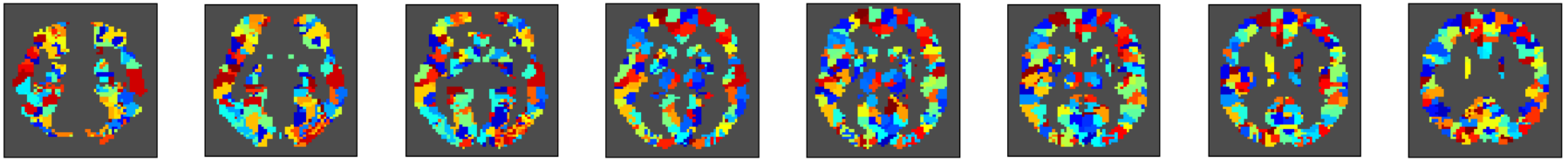
clustering from one subject

- 100s-1000 clusters
- most have at least 10s of voxels

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster searchlights into patches

SIEMENS



clustering from one subject

- 100s-1000 clusters
- most have at least 10s of voxels

each cluster yields a representative binary matrix

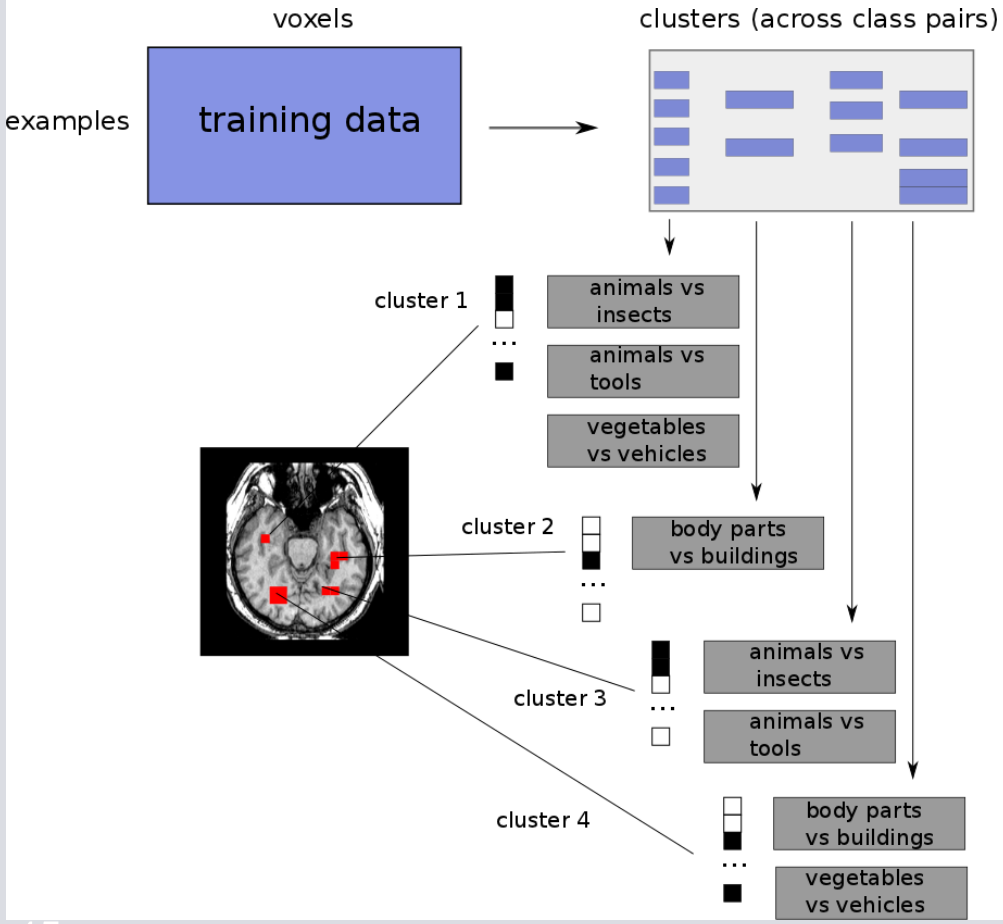
- combines distinctions made anywhere inside the cluster
- a distinction is present if it appears on more than FDR% of vox

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

build cluster pattern detectors

SIEMENS

if we view clusters as data-driven searchlights, we can train classifiers for all pairs supported: pattern detectors

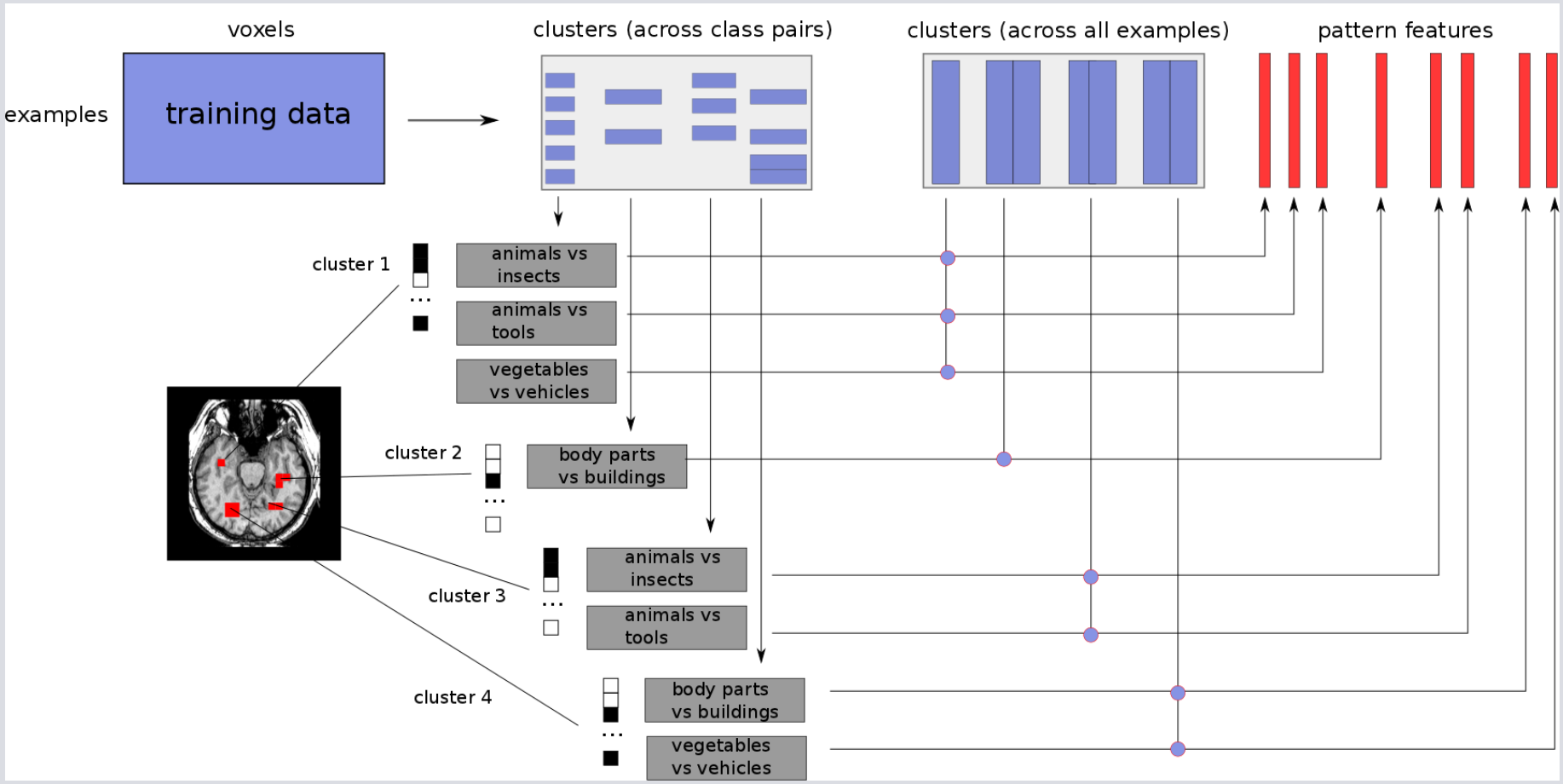


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

generate pattern features

SIEMENS

and apply those detectors to the entire training set, with their output becoming pattern features

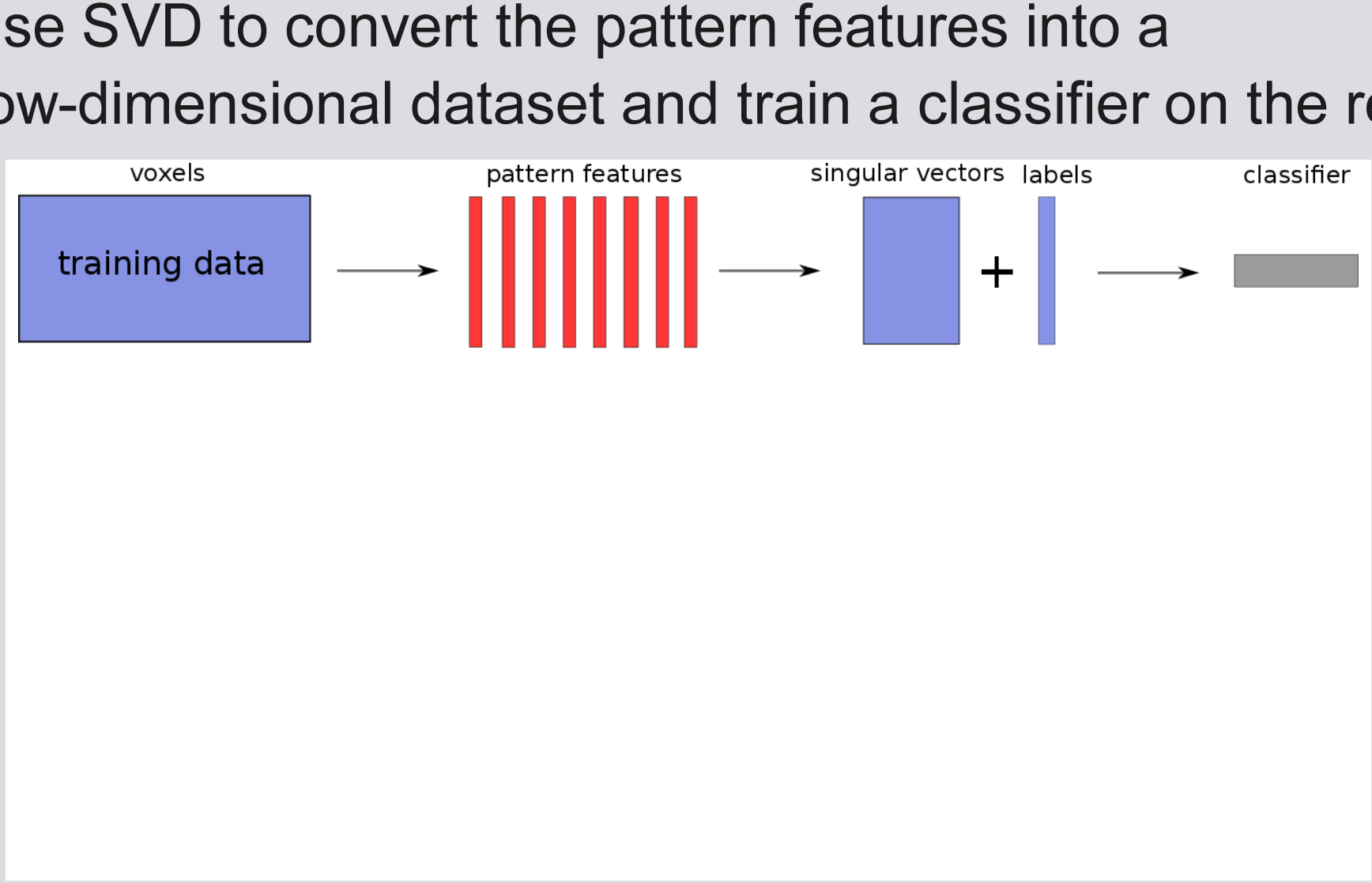


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

train pattern feature classifier

SIEMENS

use SVD to convert the pattern features into a low-dimensional dataset and train a classifier on the resu

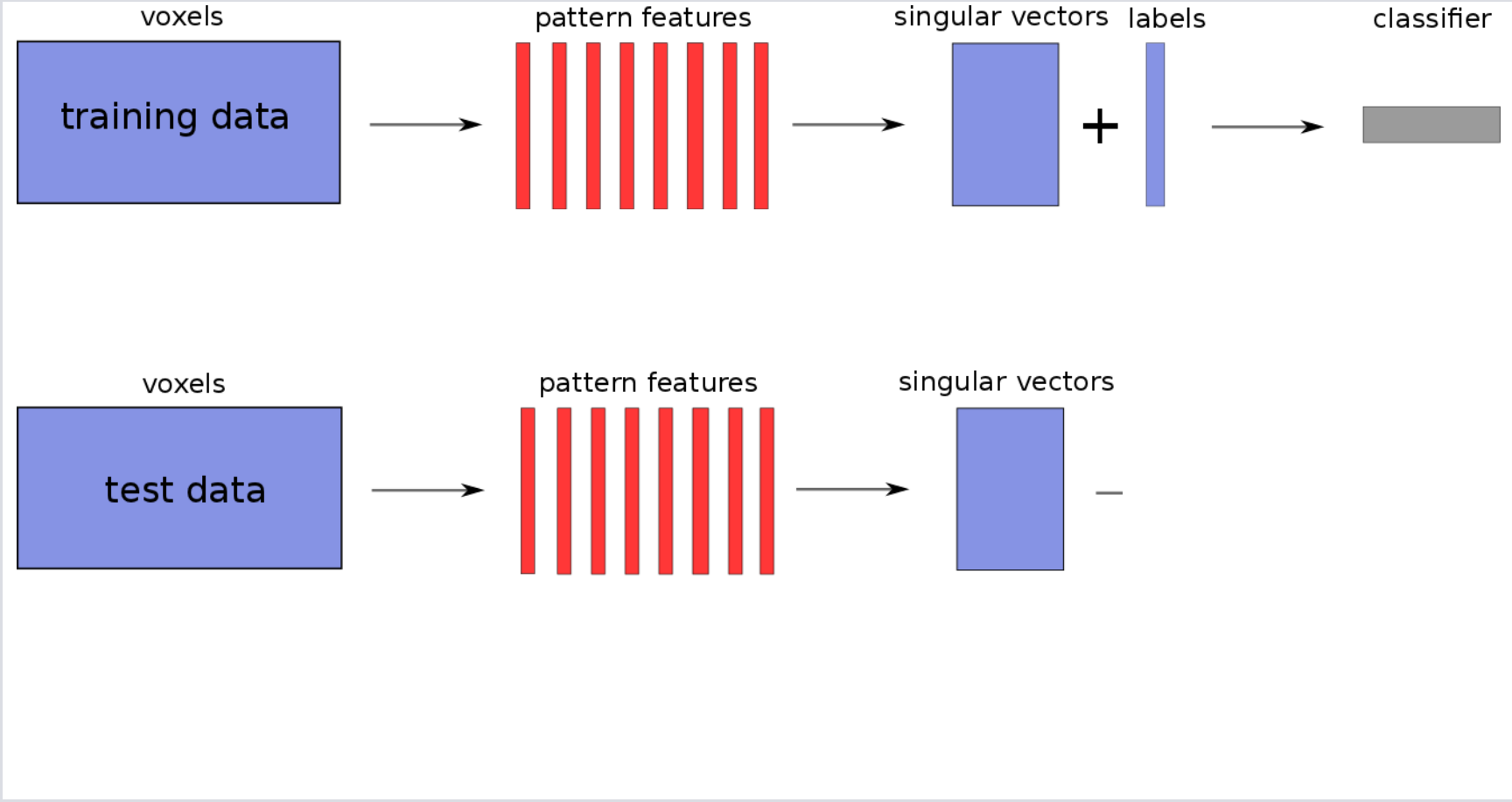


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

use pattern detectors on test set

SIEMENS

apply pattern detectors learned from the training set to the same locations on the test set, and same SVD

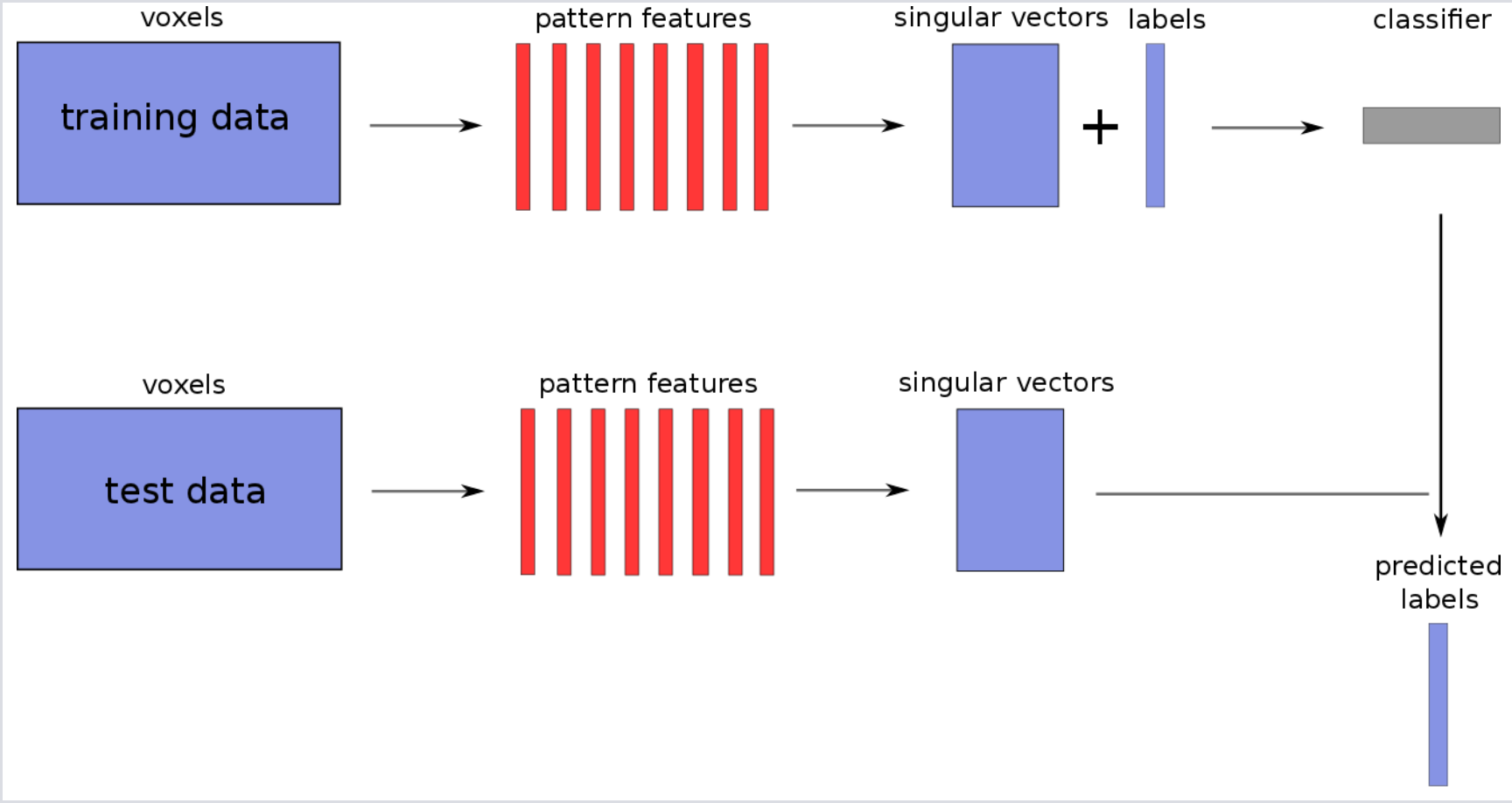


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

apply the classifier to test set

SIEMENS

and generate a label prediction



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

experimental results

SIEMENS

the data

- overall 2-fold even/odd epoch cross-validation
- training set: 180 examples = 15 x 12 classes
- searchlight cross-validation is nested in this

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

experimental results

SIEMENS

the data

- overall 2-fold even/odd epoch cross-validation
- training set: 180 examples = 15 x 12 classes
- searchlight cross-validation is nested in this

the classifiers

- LDA for individual searchlights
- linear SVMs for pattern detectors and meta
- all 180 dimensions from SVD are used

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

experimental results

SIEMENS

the baseline

- cross-validated searchlight voxel selection
- #voxels determined with nested cross-validation
- also SVD on original data (not shown)

	P1	P2	P3	P4	P5	P6	P7	P8	P9
baseline (using all voxels)	0.31	0.21	0.19	0.27	0.13	0.09	0.14	0.13	0.15
baseline (voxel selection)	0.53	0.33	0.24	0.34	0.14	0.16	0.21	0.20	0.15
#voxels selected (fold 1)	1200	400	200	1600	800	800	800	400	2000
#voxels selected (fold 2)	800	200	100	800	50	8000	100	1200	100
our method	0.54	0.34	0.32	0.43	0.15	0.19	0.22	0.20	0.18

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

interpreting the classifier

SIEMENS

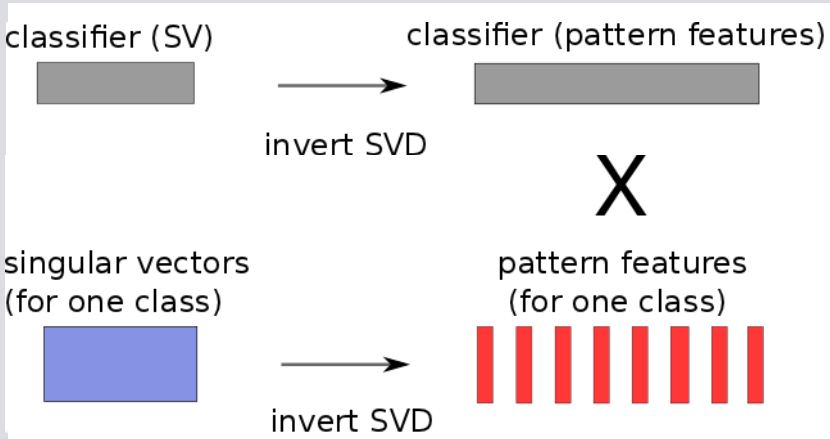
each class has a one-vs-rest linear classifier in SVD spa
with an equivalent one in the pattern feature space

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

interpreting the classifier

SIEMENS

to gauge the impact of each pattern feature in the classification we need to consider the class examples

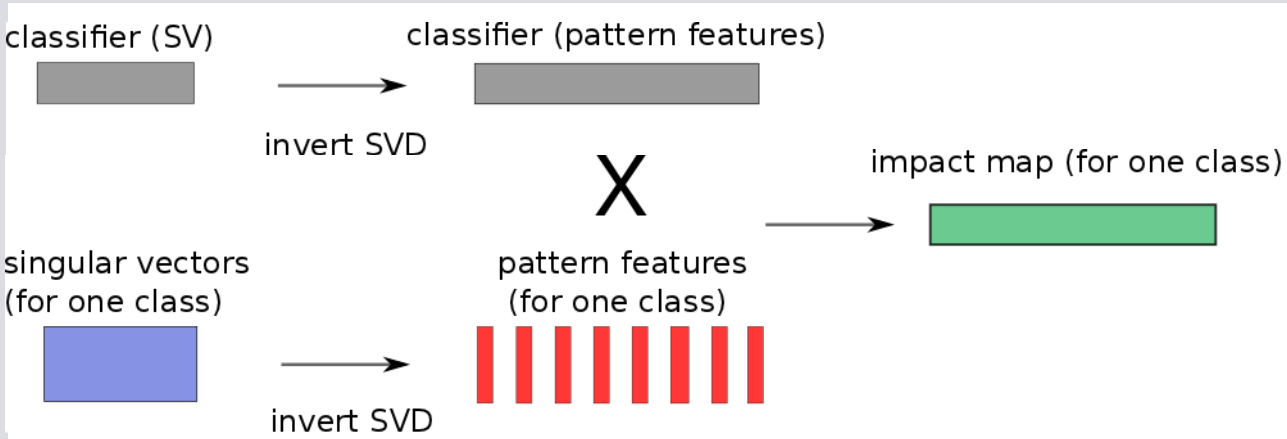


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

interpreting the classifier

SIEMENS

multiplying the one-vs-rest classifier weights by the examples of that class yields an impact map

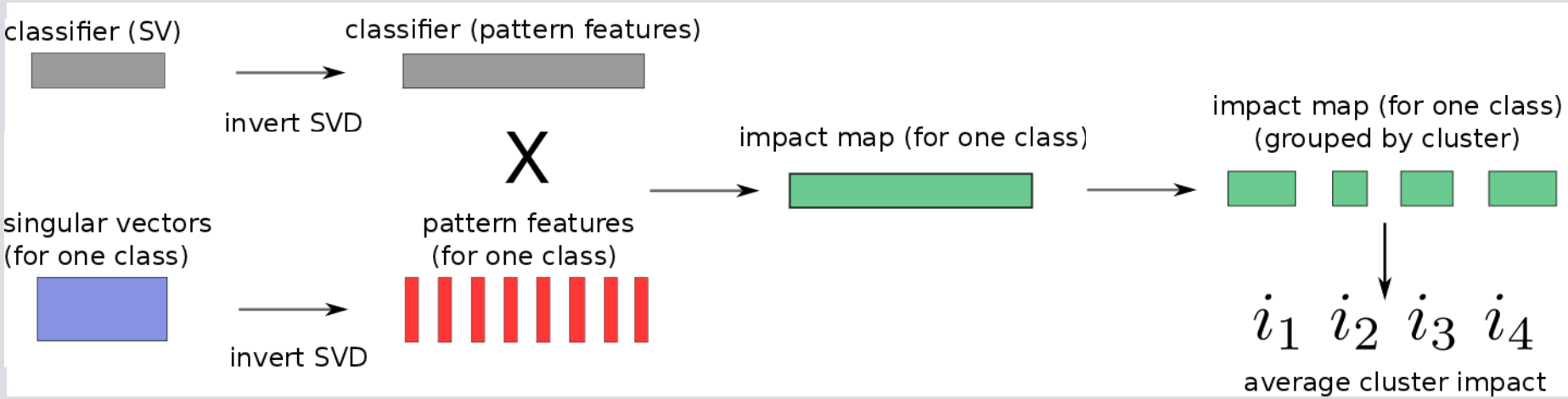


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

interpreting the classifier

SIEMENS

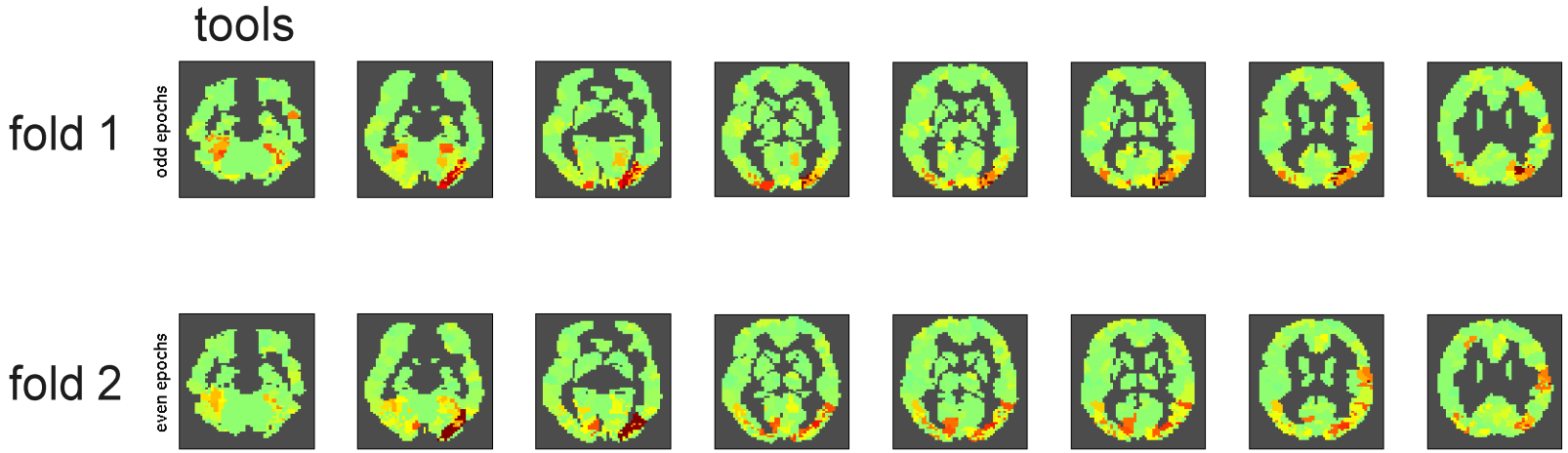
the impacts of features in each cluster can then be averaged into a single impact score for that cluster



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster impact values

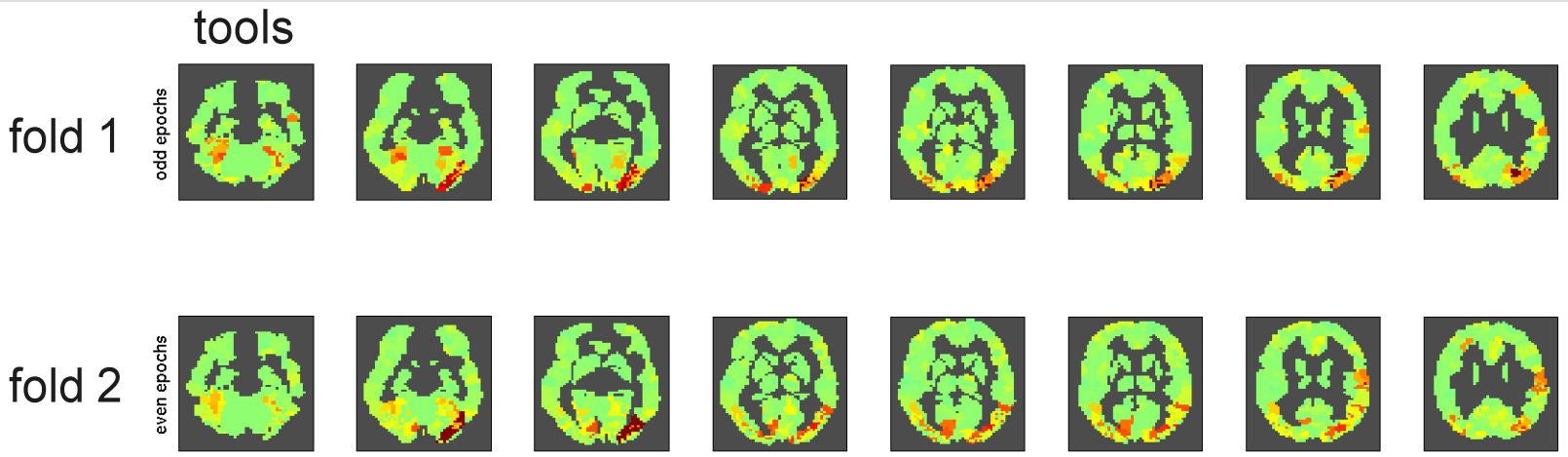
SIEMENS

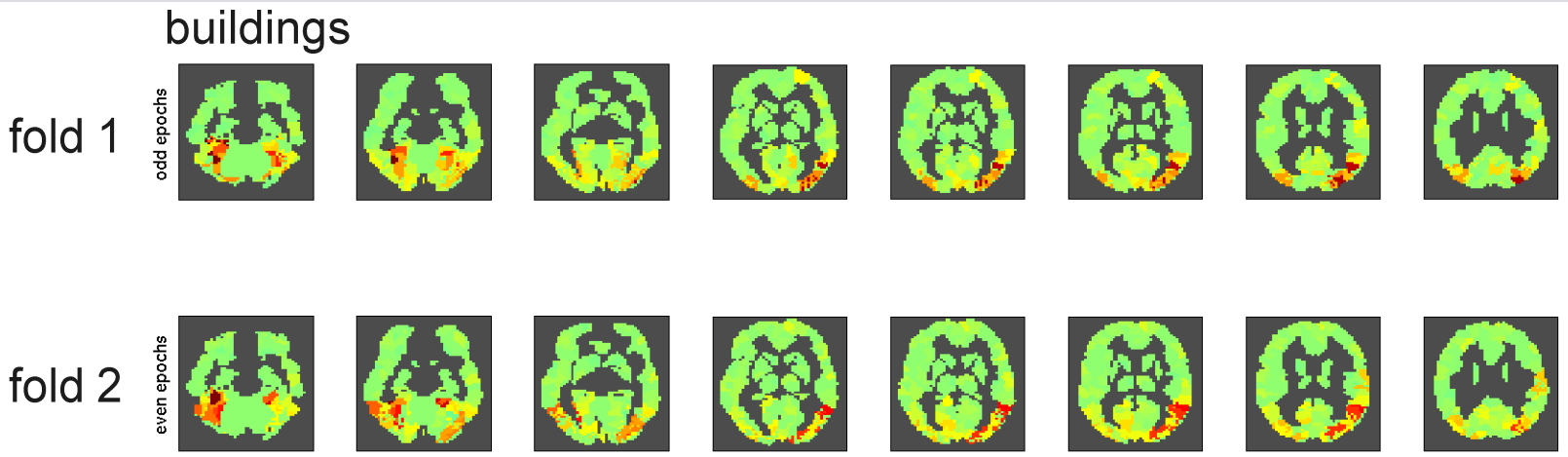


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

cluster impact values

SIEMENS





Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

conclusions

SIEMENS

- this method can perform as well or better than FS
- assign impact to localized patterns, rather than voxels
- one parameter: FDR of significant pairwise accuracy

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

conclusions

SIEMENS

- this method can perform as well or better than FS
- assign impact to localized patterns, rather than voxels
- one parameter: FDR of significant pairwise accuracy

further work:

- how to report cluster impact across multiple classes?
- what is a fair measure of reproducibility?
- how to draw group conclusions

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

outline

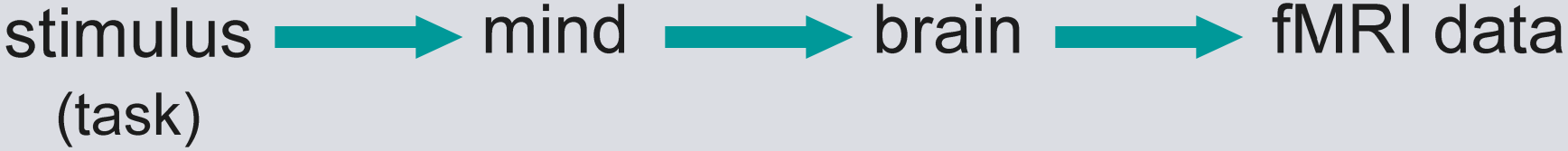
SIEMENS

- machine learning classifiers in fMRI
- a method for systematically extracting information across the brain for use in a classifier
- using text data and matrix factorization to find representations of semantic information in fMRI

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where are we?

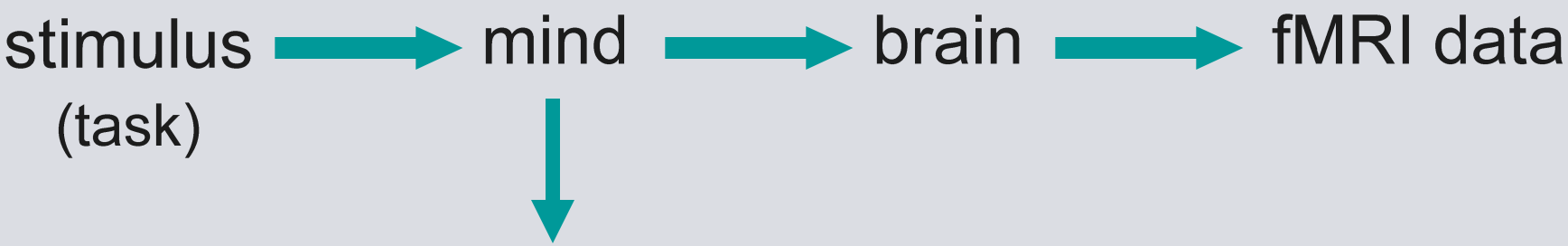
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

what do we want?

SIEMENS

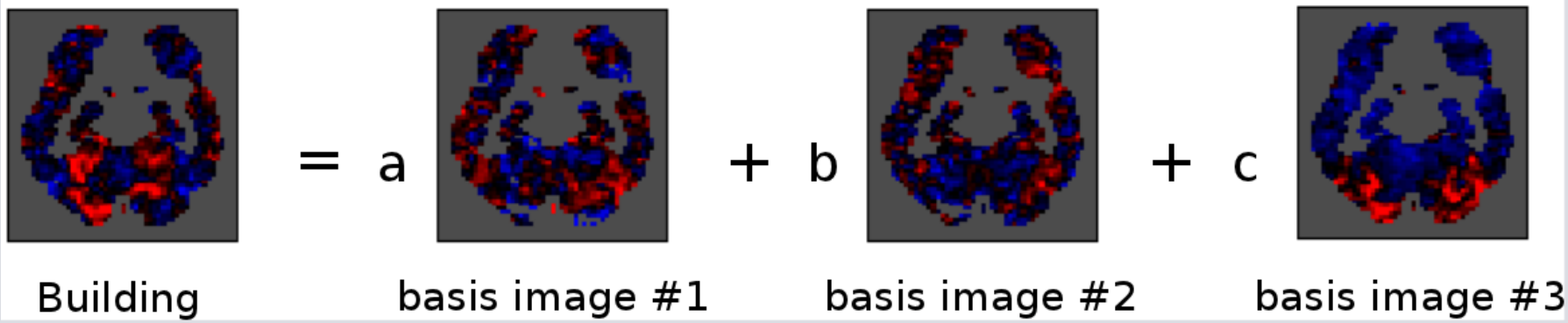


what is present in
the mind as the task
is performed?

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

image decomposition

SIEMENS

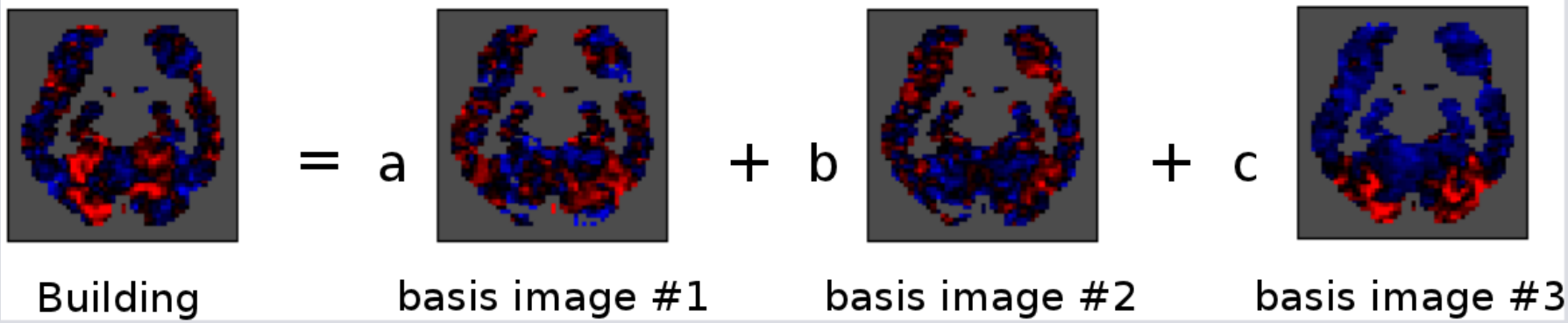


- (a,b,c) is the representation of example “Building”
- in a basis of 3 (eigen)images

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

image decomposition

SIEMENS



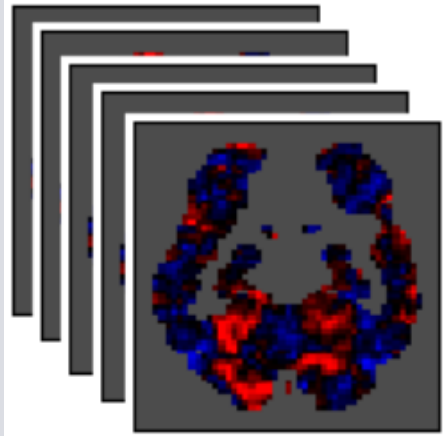
- (a,b,c) is the representation of example “Building”
- in a basis of 3 (eigen)images

- (a,b,c) can also be seen as feature values in a low-dimensional feature representation

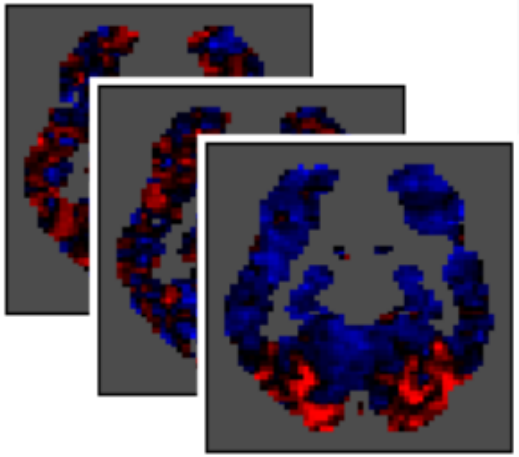
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

dataset decomposition

SIEMENS



$$= \begin{matrix} (a_1, b_1, c_1) \\ \cdots \\ (a_n, b_n, c_n) \end{matrix}$$



examples 1 to n

3 basis images

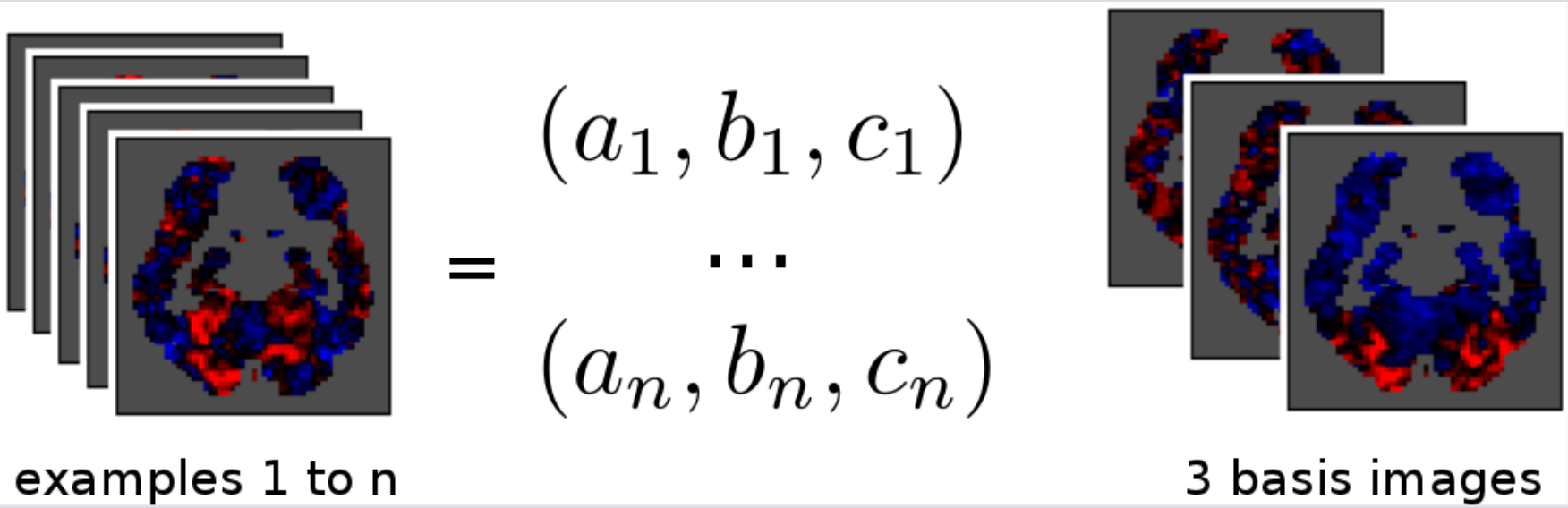
why express a dataset in terms of a basis of images?

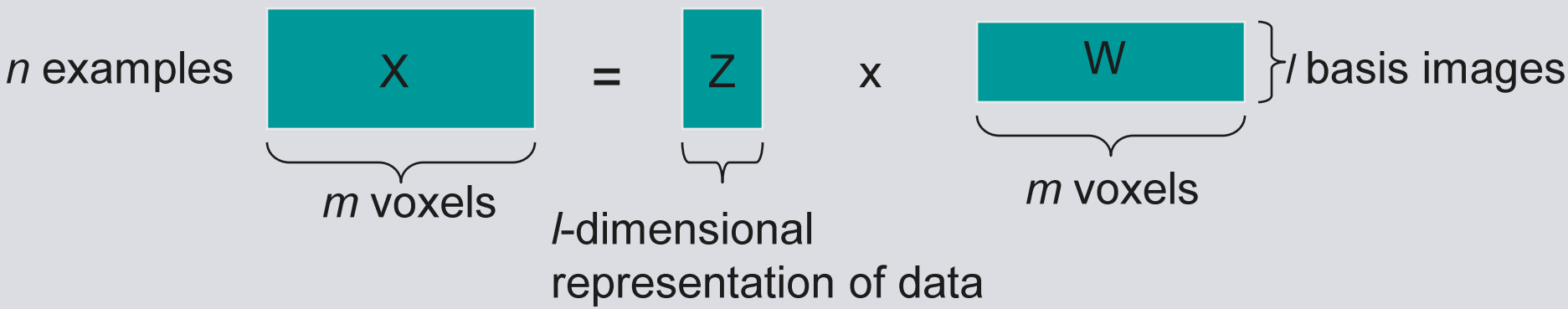
- capture spatial patterns of activity over many voxels
- coordinates give a succinct representation of the data

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

dataset decomposition

SIEMENS

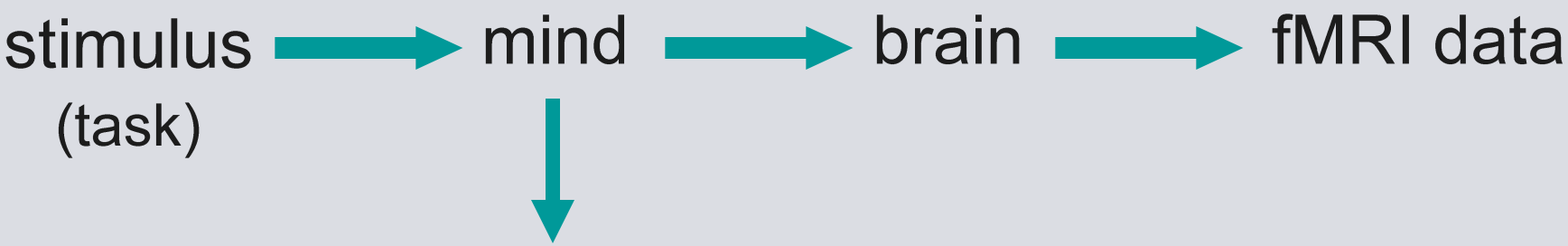




Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

what do we want?

SIEMENS

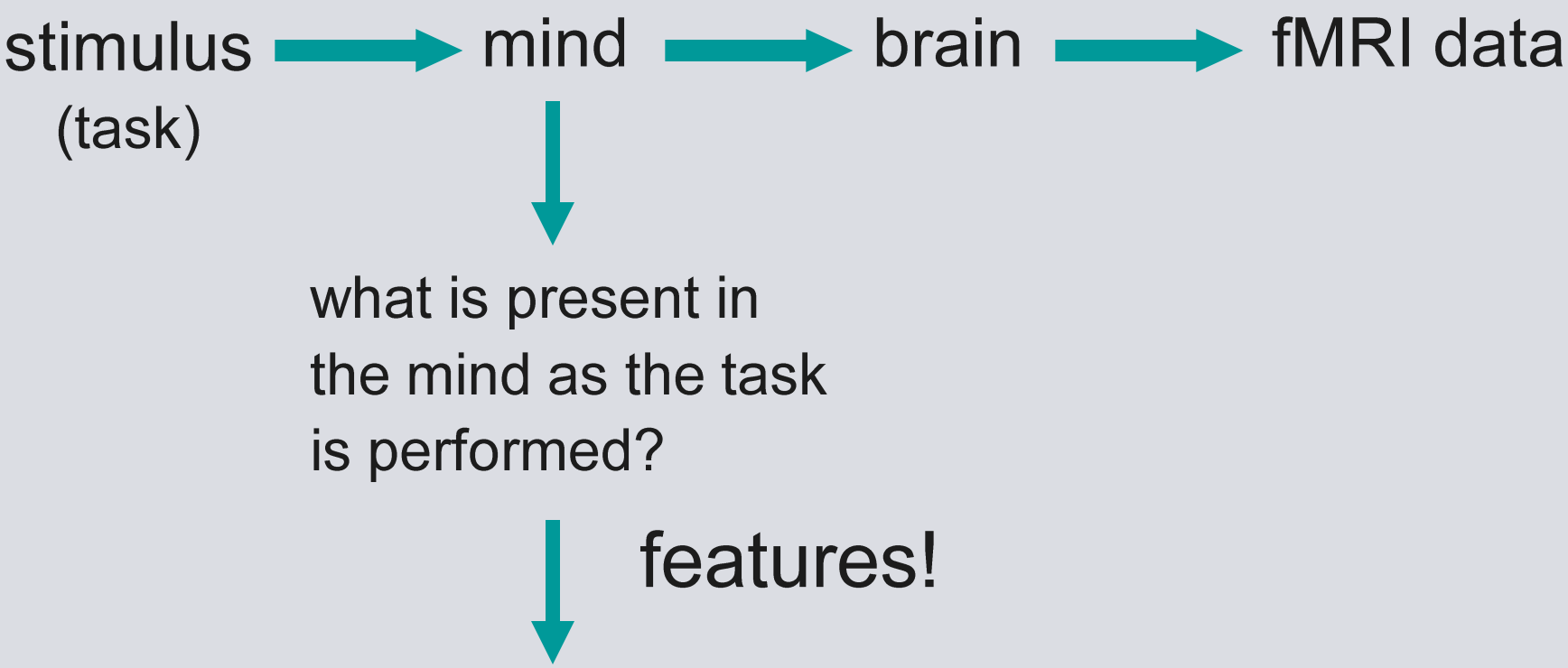


what is present in
the mind as the task
is performed?

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where can it come from?

SIEMENS



- known or constrained from behavioural experiments
- modelled mathematically or computationally
- hypothesized
- learnt elsewhere

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where do features come from?

SIEMENS

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters
  - supervised matrix factorization (SVDM, SMF)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

SVDM?

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where do features come from?

SIEMENS

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters
  - supervised matrix factorization (SVDM, SMF)

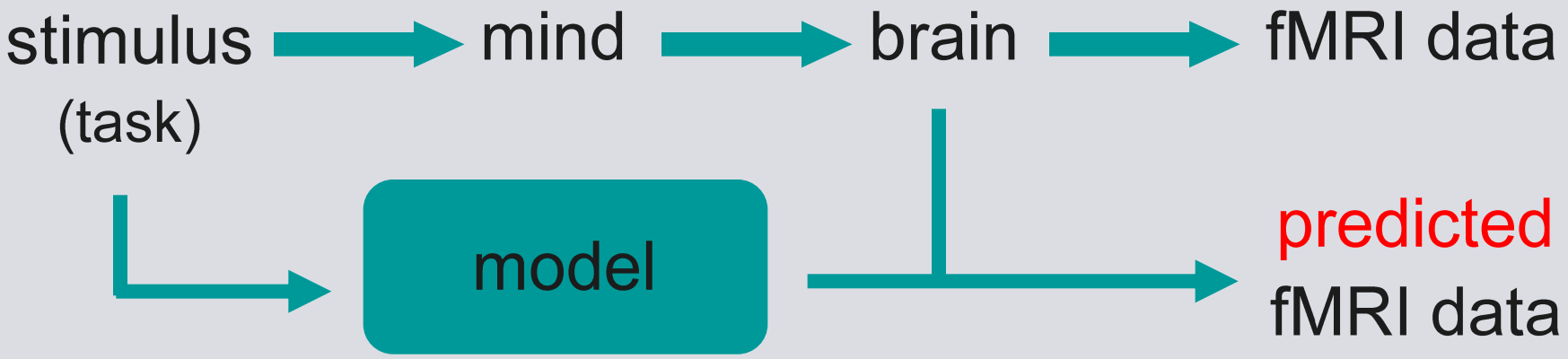
- hypothesized
  - behavioural experiments
  - mathematical model

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

how do we know they are there?

SIEMENS

1



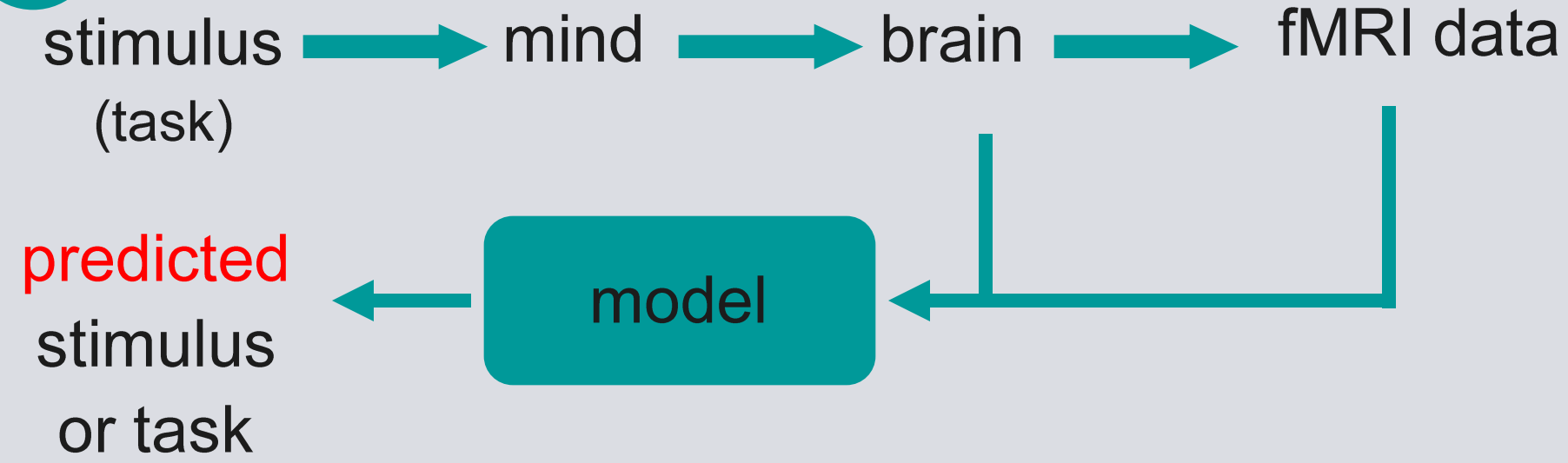
- Kay et al. 2008
- Mitchell et al. 2008

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

how do we know they are there?

SIEMENS

2



- Thirion et al. 2006
- Miyawaki et al. 2008
- Naselaris et al. 2009
- van Gerven et al. 2010
- Pereira et al. 2011
- Nishimoto et al. 2011

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where do features come from?

SIEMENS

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters
  - supervised matrix factorization (SVDM, SMF)

- hypothesized
  - behavioural experiments
  - mathematical model

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

where do features come from?

SIEMENS

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters
  - supervised matrix factorization (SVDM, SMF)

- hypothesized
  - behavioural experiments
  - mathematical model

- learned from text

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

learning features from text

SIEMENS

- from text alone [Landauer+Dumais 97, Griffiths et al 07]
  - LSA, topic models
- verbs + co-occurrence with them [Mitchell et al 08]
- text + people [Pereira et al. 2011, in press]
  - use word norms to choose concrete concepts
  - collect wikipedia articles for those concepts
  - build topic model from articles

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic models

SIEMENS

[Blei et al 2003]

- used to model a document corpus
- documents are bags-of-words: words and counts
- model:
  - per document: the probability of each topic
  - per topic: a probability distribution over words

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic models and semantic features

SIEMENS

[Pereira et al. 2011, in press]

- if a document corresponds to a concept…
  - topics ⟷ semantic features
  - topic probabilities ⟷ semantic feature values

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic models and semantic features

SIEMENS

[Pereira et al. 2011, in press]

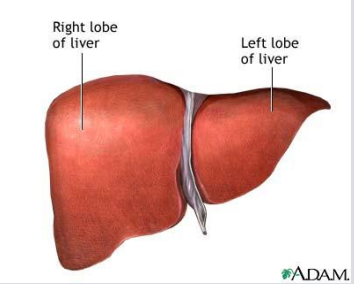
- if a document corresponds to a concept…
  - topics ⟷ semantic features
  - topic probabilities ⟷ semantic feature values

- the "Wee-kipedia"
  - corpus of 3500 Wikipedia hand-picked articles
  - articles about concrete/imageable concepts
  - vocabulary of about 50K words

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

our topic model(s)

SIEMENS

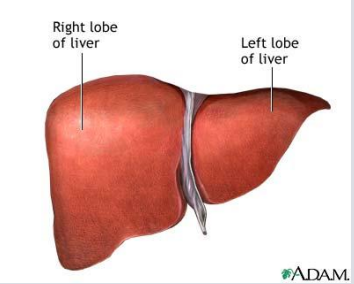


http://en.wikipedia.org/wiki/Liver

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

our topic model(s)

SIEMENS



http://en.wikipedia.org/wiki/Liver

extract basic text

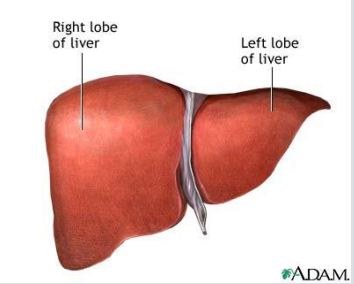
The liver is a vital organ present in vertebrates and some other animals; it has a wide range of functions, a few of which are detoxification, protein synthesis, and production of biochemicals necessary for digestion. The liver is necessary for survival; a human can only survive up to 24 hours without liver function. (...)

Mammal and bird livers are commonly eaten as food by humans. Liver can be baked, boiled, broiled, fried or eaten raw, but is perhaps most commonly made into spreads.

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

our topic model(s)

SIEMENS



http://en.wikipedia.org/wiki/Liver

↓ extract basic text

The liver is a vital organ present in vertebrates and some other animals; it has a wide range of functions, a few of which are detoxification, protein synthesis, and production of biochemicals necessary for digestion. The liver is necessary for survival; a human can only survive up to 24 hours without liver function. (...)

Mammal and bird livers are commonly eaten as food by humans. Liver can be baked, boiled, broiled, fried or eaten raw, but is perhaps most commonly made into spreads.

↓ lemmatise (remove stopwords later)

the liver be a vital organ present in vertebrate and some other animal it have a wide range of function a few of which be detoxification protein synthesis and production of biochemical necessary for digestion the liver be necessary for survival a human can only survive up to 24 hour without liver function (...)

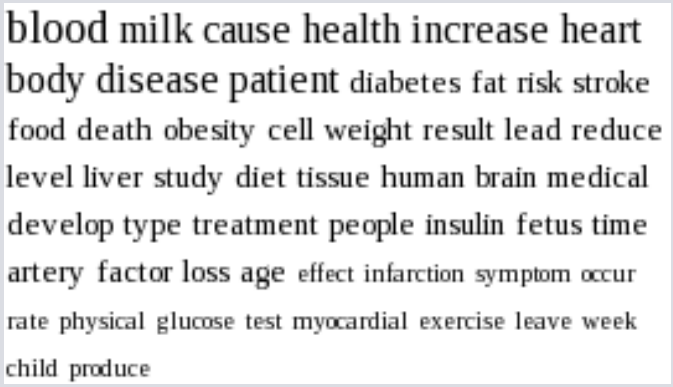
mammal and bird liver be commonly eat as food by human liver can be bake boil broil fry or eat raw but be perhap most commonly make into spread

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from concept to topic

SIEMENS

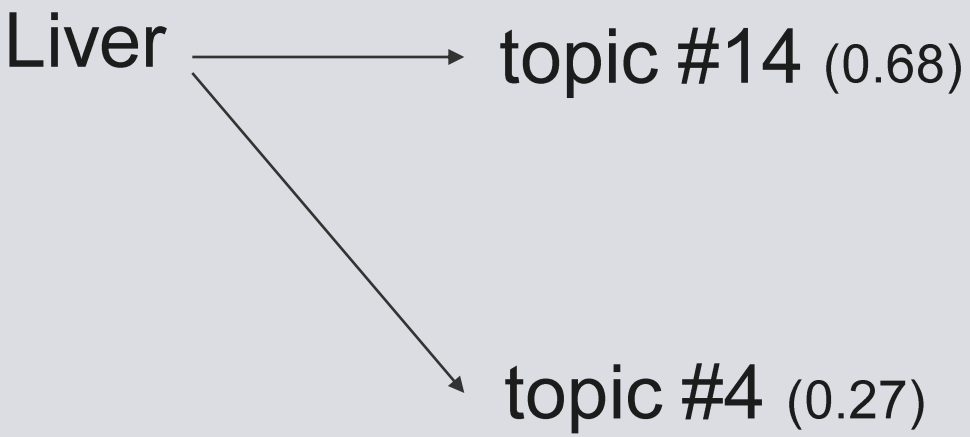
Liver → topic #14 (0.68)



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from concept to topic

SIEMENS



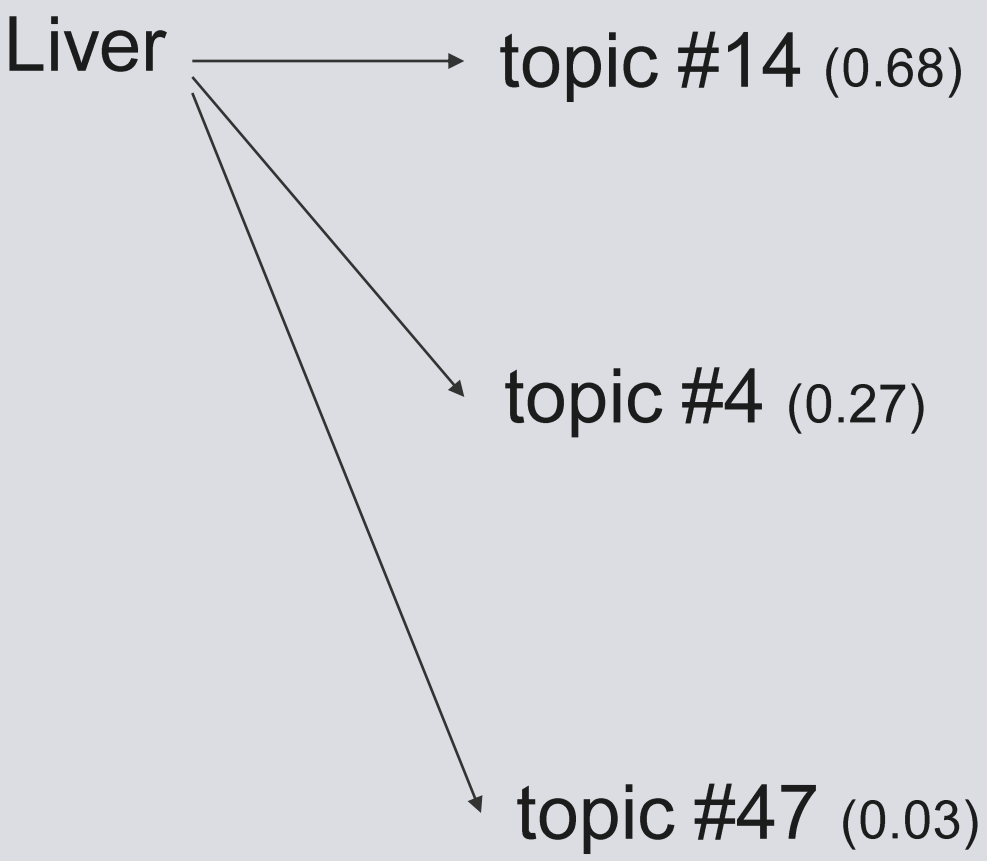
blood milk cause health increase heart body disease patient diabetes fat risk stroke food death obesity cell weight result lead reduce level liver study diet tissue human brain medical develop type treatment people insulin fetus time artery factor loss age effect infarction symptom occur rate physical glucose test myocardial exercise leave week child produce

muscle bone frog nerve body branch join limb human tissue form anatomy ligament lateral structure leg function skin cell anterior iris lower organ medial surface birch head foot spinal cord posterior movement skull contain knee lung artery gland position cactus animal region hip lip tendon upper attach vein refer locate

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from concept to topic

SIEMENS



blood milk cause health increase heart body disease patient diabetes fat risk stroke food death obesity cell weight result lead reduce level liver study diet tissue human brain medical develop type treatment people insulin fetus time artery factor loss age effect infarction symptom occur rate physical glucose test myocardial exercise leave week child produce

muscle bone frog nerve body branch join limb human tissue form anatomy ligament lateral structure leg function skin cell anterior iris lower organ medial surface birch head foot spinal cord posterior movement skull contain knee lung artery gland position cactus animal region hip lip tendon upper attach vein refer locate

food fruit cook sugar eat potato meat rice pepper variety dish serve sauce vegetable butter cuisine juice sweet bread flavor popular ingredient milk breakfast contain spice common meal dry product fry produce add word hot fresh refer Unite traditional vinegar egg sausage water salt taste soup state type country

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from topic to concept

SIEMENS

topic #14 (0.68)

Diabetes_mellitus (0.993) Stroke (0.991) Myocardial_infarction (0.990) Fetus (0.914) Overweight (0.898) Amniotic_fluid (0.879) Obesity (0.869) Blood_vessel (0.843) Hunger (0.833) Umbilical_cord (0.782) Heart (0.756) Starvation (0.677) Liver (0.676) Autopsy (0.641) Gold_standard_(test) (0.636) Milk (0.627) Health (0.618) Blood_bank (0.614) Malnutrition (0.584) Adipose_tissue (0.559) Dietary_fiber (0.540) Blood_type (0.496) Death (0.495) Portal_venous_system (0.481) Diet_food (0.462) Hyperthermia (0.428) Fetus_(biology) (0.422) Udder (0.412) Blood (0.408) Death_rattle (0.397)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from topic to concept

SIEMENS

topic #14 (0.68)

Diabetes_mellitus (0.993) Stroke (0.991) Myocardial_infarction (0.990) Fetus (0.914) Overweight (0.898) Amniotic_fluid (0.879) Obesity (0.869) Blood_vessel (0.843) Hunger (0.833) Umbilical_cord (0.782) Heart (0.756) Starvation (0.677) Liver (0.676) Autopsy (0.641) Gold_standard_(test) (0.636) Milk (0.627) Health (0.618) Blood_bank (0.614) Malnutrition (0.584) Adipose_tissue (0.559) Dietary_fiber (0.540) Blood_type (0.496) Death (0.495) Portal_venous_system (0.481) Diet_food (0.462) Hyperthermia (0.428) Fetus_(biology) (0.422) Udder (0.412) Blood (0.408) Death_rattle (0.397)

topic #4 (0.27)

Human_leg (0.990) Spinal_cord (0.978) Knee (0.974) Elbow (0.944) Arm (0.943) Hip (0.931) Human_abdomen (0.925) Shoulder (0.923) Ankle (0.904) Neck (0.862) Wrist (0.849) Bone (0.771) Thigh (0.761) Hand (0.750) Chest (0.684) Abdomen (0.671) Hip_bone (0.665) Neural_pathway (0.621) Torso (0.607) Lung (0.600) Toe (0.569) Sclera (0.565) Lip (0.563) Throat (0.560) Limb_(anatomy) (0.549) Nerve (0.545) Mouth (0.541) Skull (0.501) Leaf (0.490) Thorn_(botany) (0.487)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

from topic to concept

SIEMENS

topic #14 (0.68)

Diabetes_mellitus (0.993) Stroke (0.991) Myocardial_infarction (0.990) Fetus (0.914) Overweight (0.898) Amniotic_fluid (0.879) Obesity (0.869) Blood_vessel (0.843) Hunger (0.833) Umbilical_cord (0.782) Heart (0.756) Starvation (0.677) Liver (0.676) Autopsy (0.641) Gold_standard_(test) (0.636) Milk (0.627) Health (0.618) Blood_bank (0.614) Malnutrition (0.584) Adipose_tissue (0.559) Dietary_fiber (0.540) Blood_type (0.496) Death (0.495) Portal_venous_system (0.481) Diet_food (0.462) Hyperthermia (0.428) Fetus_(biology) (0.422) Udder (0.412) Blood (0.408) Death_rattle (0.397)

topic #4 (0.27)

Human_leg (0.990) Spinal_cord (0.978) Knee (0.974) Elbow (0.944) Arm (0.943) Hip (0.931) Human_abdomen (0.925) Shoulder (0.923) Ankle (0.904) Neck (0.862) Wrist (0.849) Bone (0.771) Thigh (0.761) Hand (0.750) Chest (0.684) Abdomen (0.671) Hip_bone (0.665) Neural_pathway (0.621) Torso (0.607) Lung (0.600) Toe (0.569) Sclera (0.565) Lip (0.563) Throat (0.560) Limb_(anatomy) (0.549) Nerve (0.545) Mouth (0.541) Skull (0.501) Leaf (0.490) Thorn_(botany) (0.487)

topic #47 (0.03)

Breakfast (0.992) Curry (0.980) French_toast (0.957) Pickling (0.947) Mustard_(condiment) (0.943) Pickled_cucumber (0.933) Coconut_milk (0.924) Supper (0.920) Soup (0.905) Fruit_preserves (0.902) Salad (0.884) Sauce (0.869) Pudding (0.864) Sichuan_pepper (0.858) Brunch (0.826) Capsicum (0.817) Butter (0.814) Carrot_cake (0.802) Side_dish (0.795) Sponge_cake (0.787) Lemon (0.787) Condiment (0.782) List_of_cuisines (0.765) Lime_(fruit) (0.765) Sausage (0.763) Peel_(fruit) (0.757) Black_pudding (0.756) Orange_juice (0.752) Macaroni (0.746) Macaroni_and_cheese (0.744)

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic model

SIEMENS

K-topic representation of each article

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic model

SIEMENS

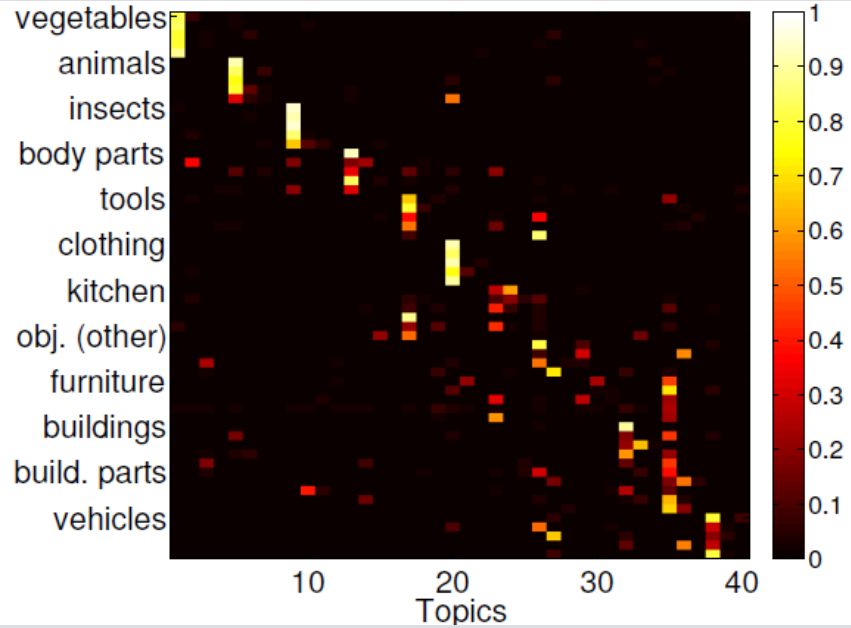
Table
Hammer
Bear
...

wiki = Z and word distribution } K topics

40K words

40K words

K-topic representation of each



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

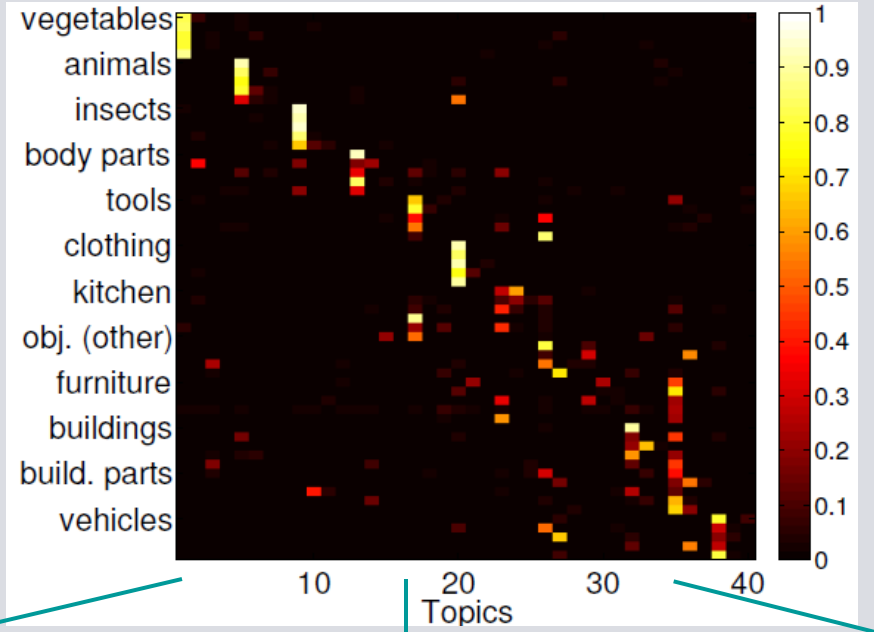
topic model

SIEMENS

Table
Hammer
Bear
...

wiki (40K words) = Z and word distribution (40K words) } K topics

K-topic representation of each



plant fruit seed grow leaf flower tree sugar produce species

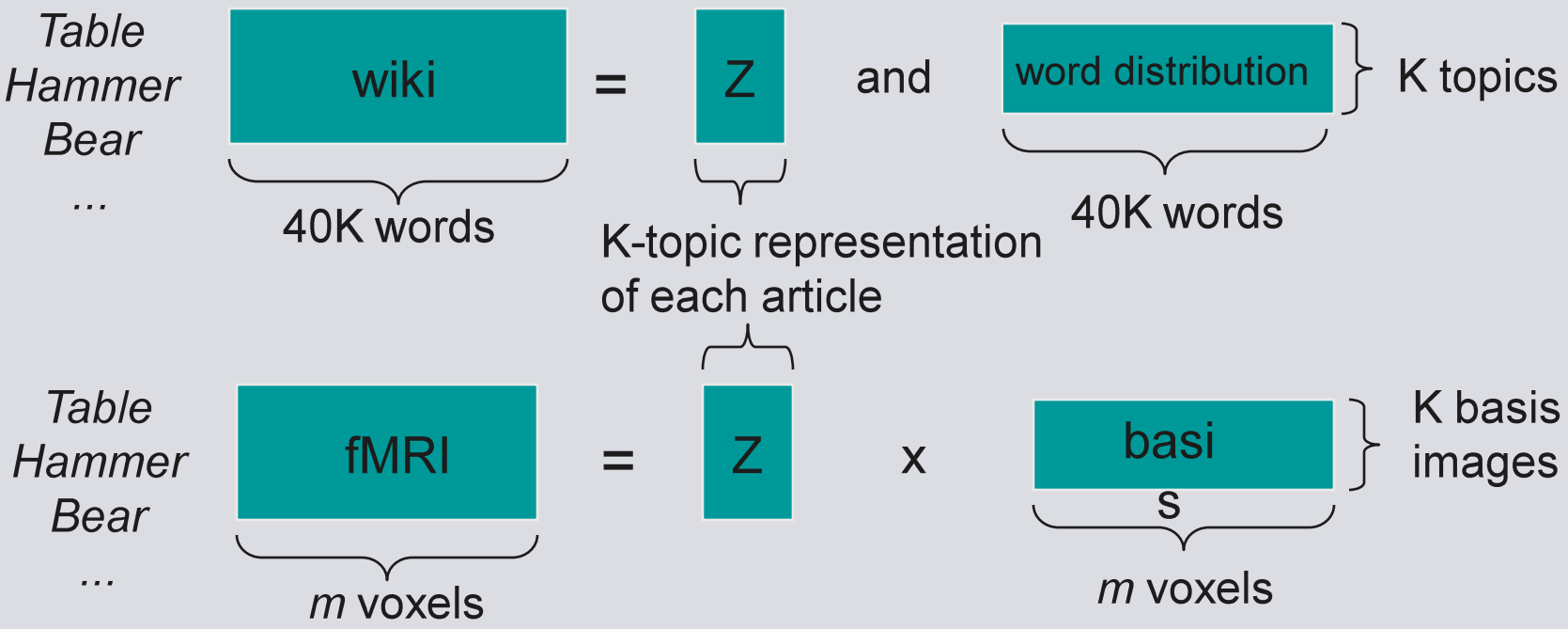
iron blade steel handle head cut hair metal tool nail

material wood paint build wall structure construction design size window

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

topic model and fMRI model

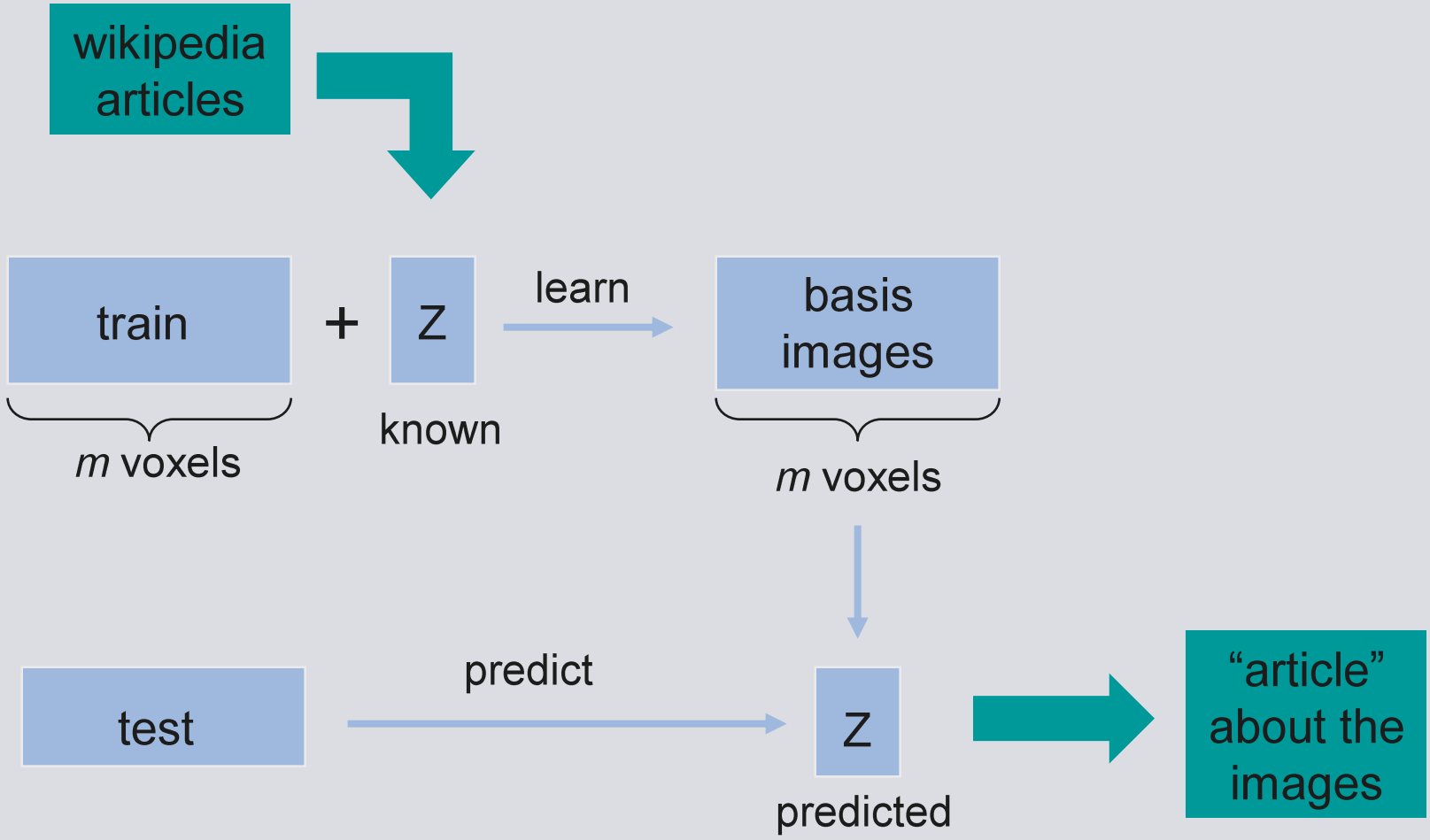
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

predicting text

SIEMENS

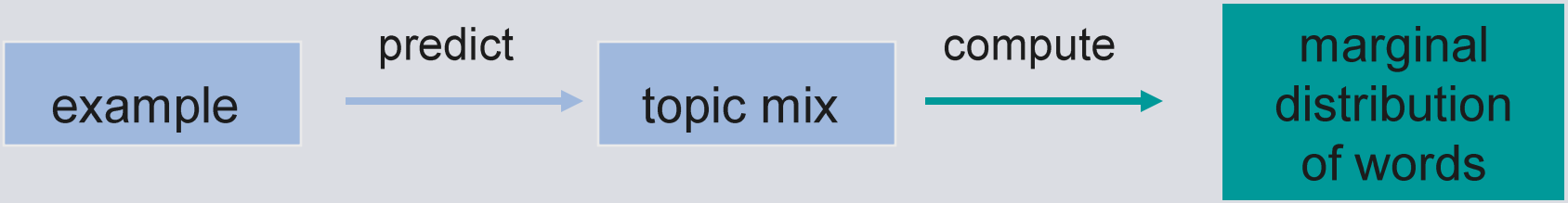


Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

predicting text

SIEMENS

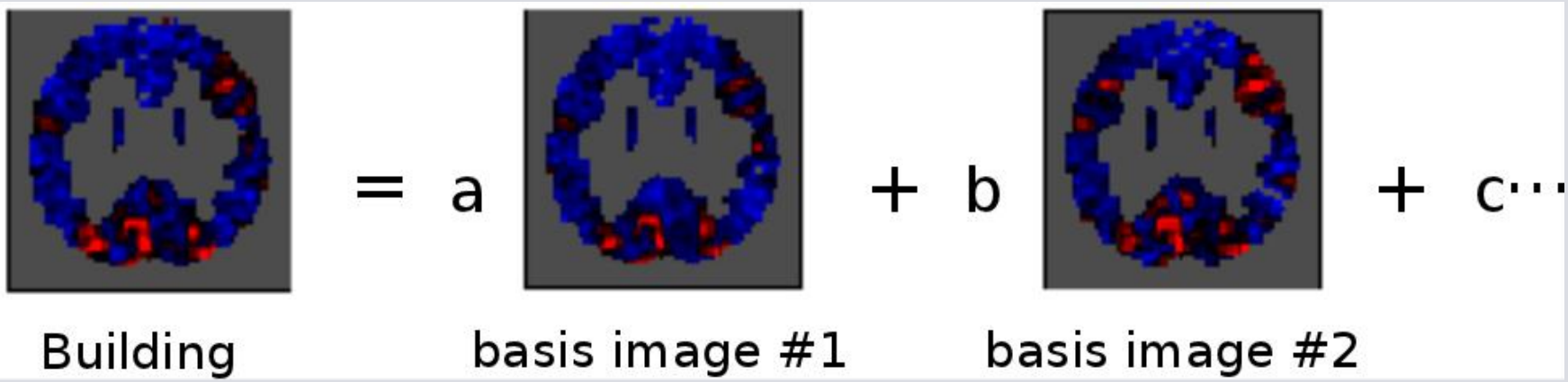
- original article is a “bag of words”
- a topic is a distribution over words
- article comes from a mixture of topics



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

predicting text

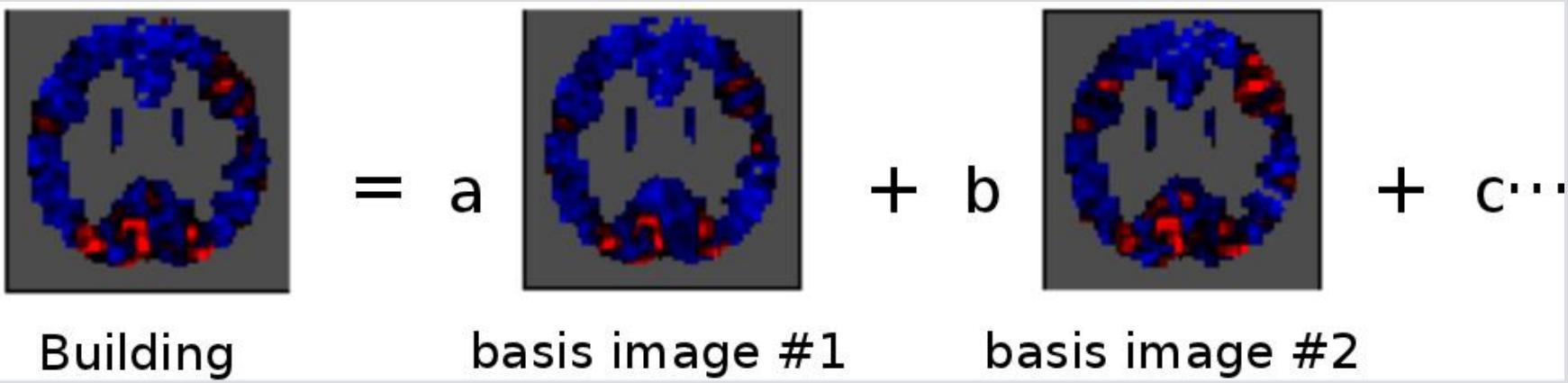
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

predicting text

SIEMENS



city	material	build		city	house	street	material	wood	build
wood	street	town	wall	= a	build	town	home	+ b	wall	design	window	paint
window	home	store		state	store	bus	road	size	structure	construction

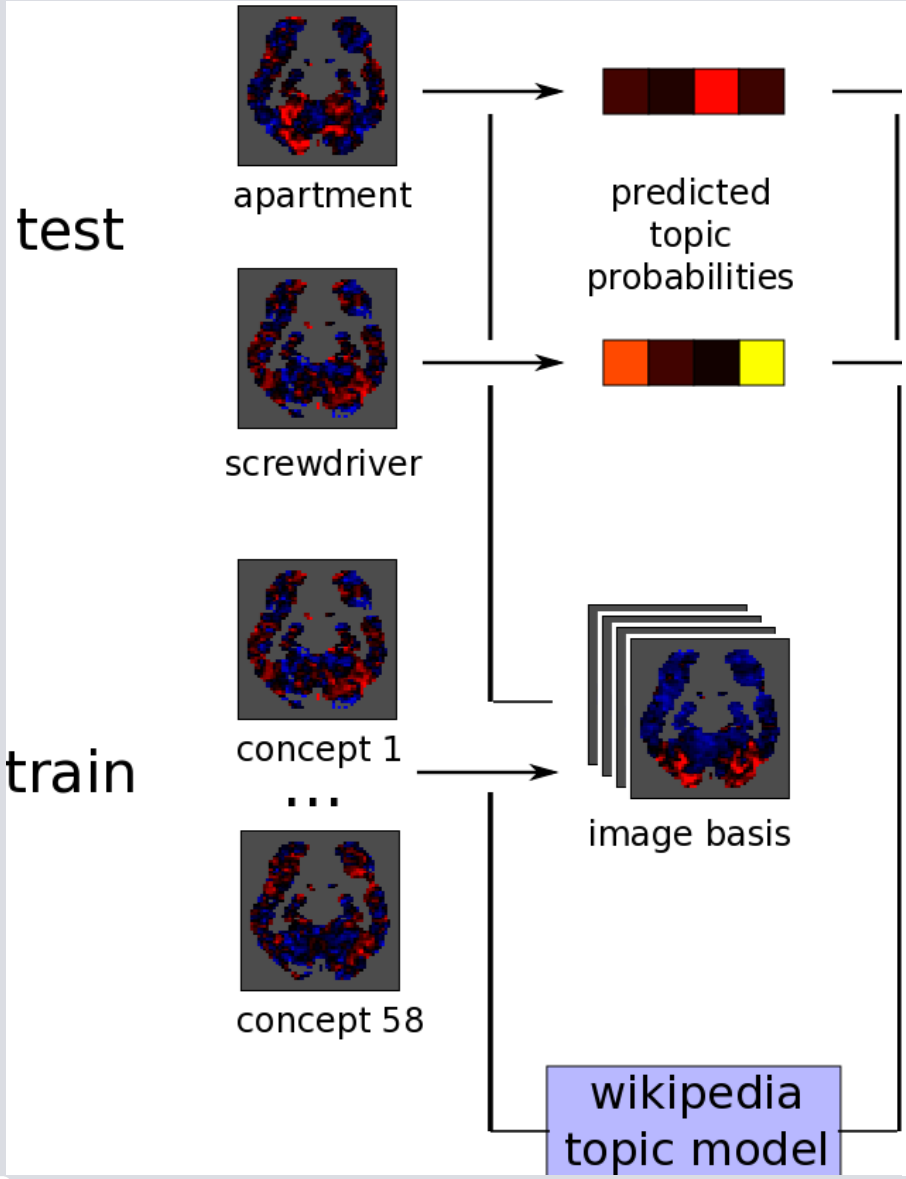
online model browser:

http://minerva.csbmb.princeton.edu/wikipedia

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classification via word probability

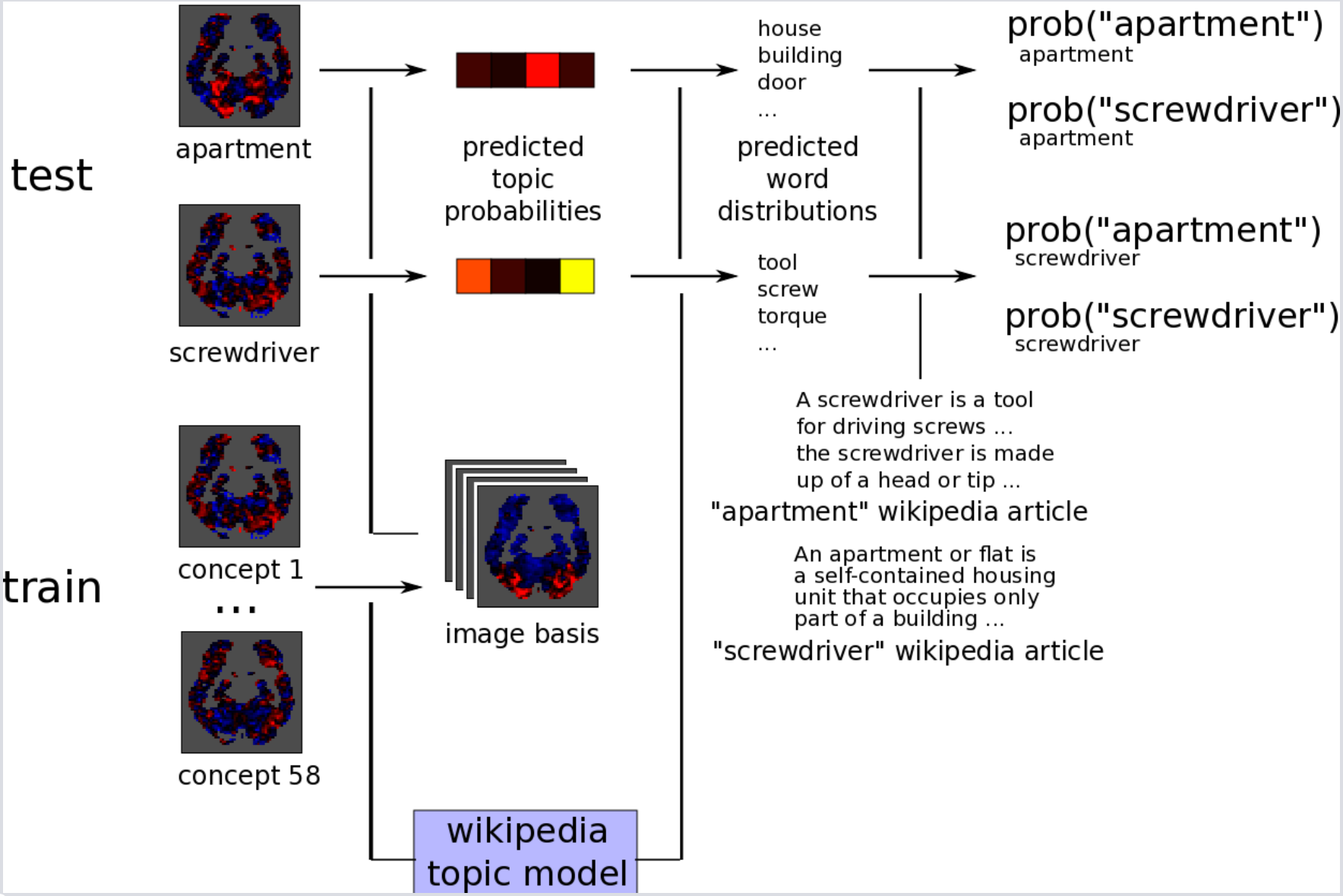
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classification via word probability

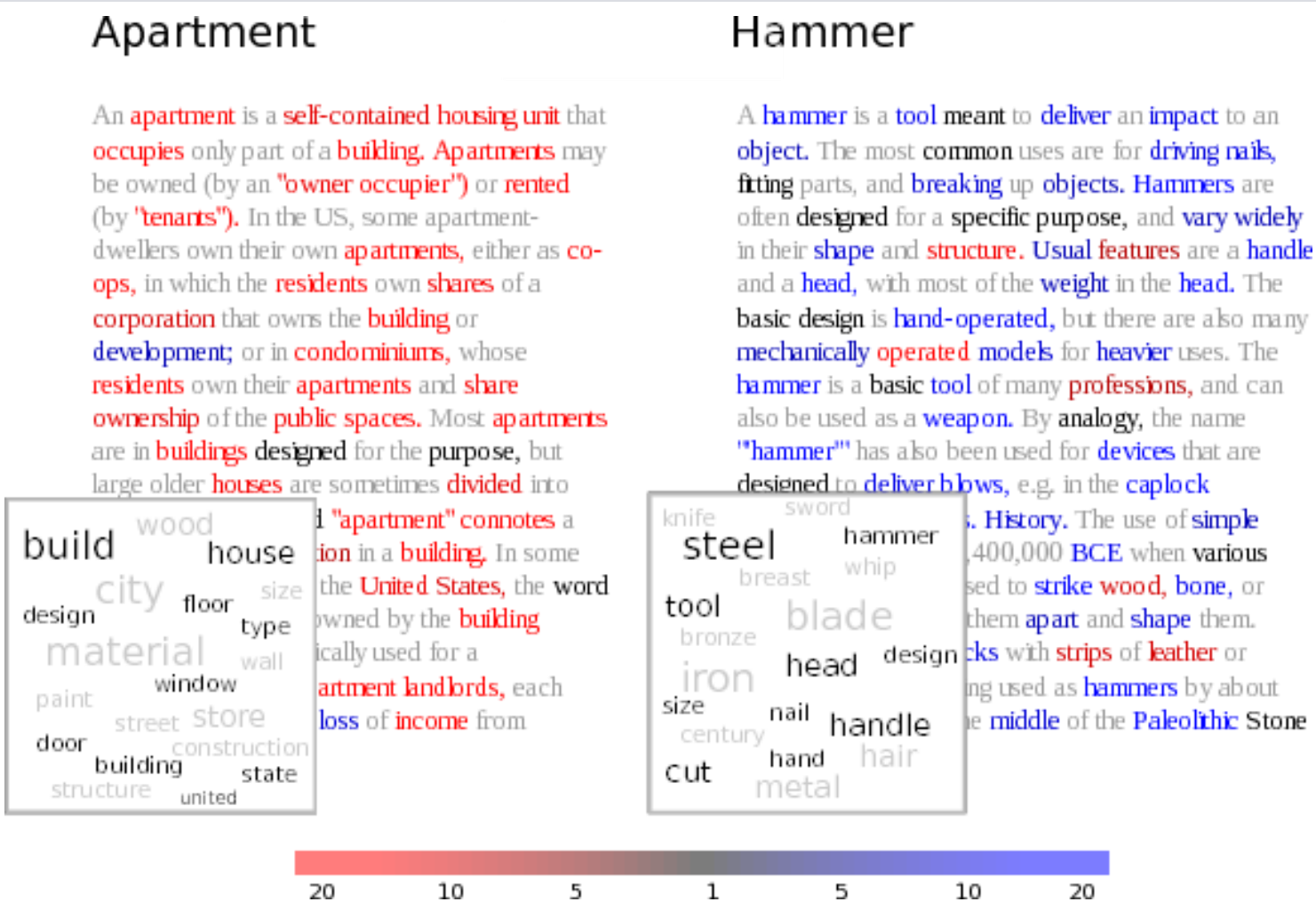
SIEMENS



Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classification via word probability

SIEMENS

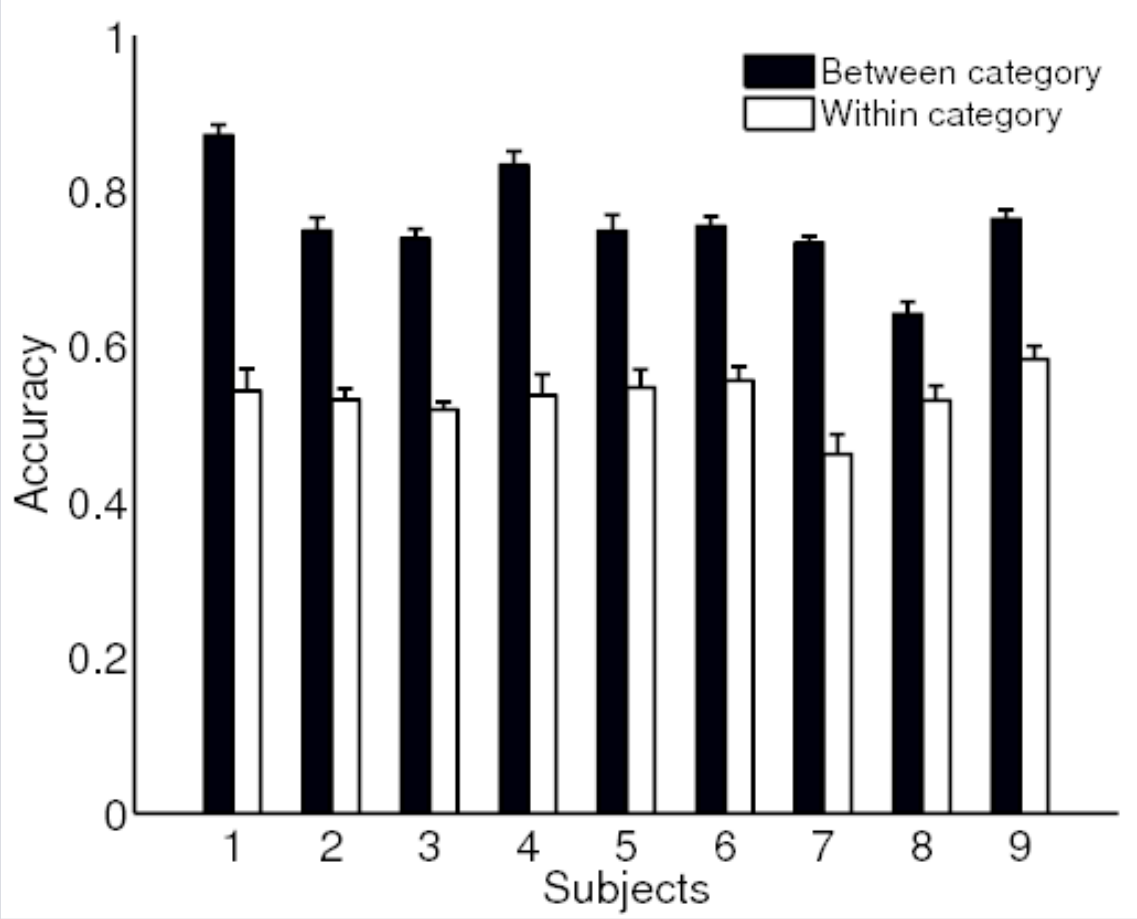


probability ratio assigned to each word in brain-derived distributions

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

classification via word probability

SIEMENS

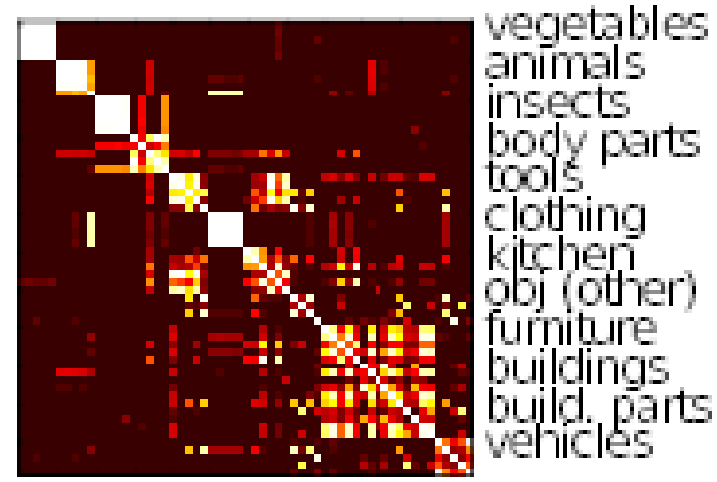


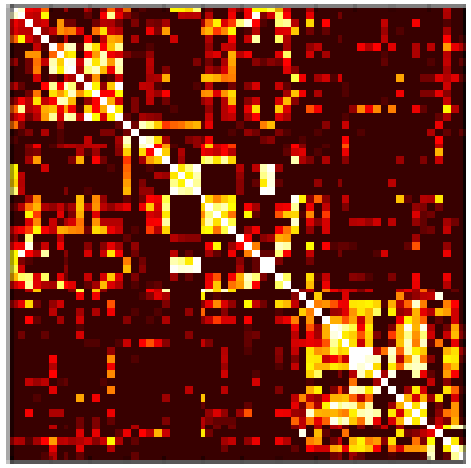
results averaged over several topic models to prevent bias

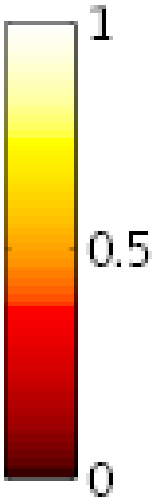
Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

representation similarity

SIEMENS







Topic probabilities (from text)

Topic probabilities (from fMRI data)

results from best subject

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

conclusions

SIEMENS

- we can learn features directly from text
- learned features capture semantic knowledge
- they yield better decompositions than alternatives
- further work
  - finer grained topic models/basic features
  - new experiment involving reading

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.

Thank you!

Copyright © Siemens Corporation, Corporate Research & Technology, Inc.