A New Challenge of Information Processing in the Web Age

Bo Zhang
Tsinghua National Lab for Information Science & Technology
State Key Lab of Intelligent Technology & Systems
Information Science & Technology College
Computer Science & Technology Department
Tsinghua University, Beijing China

Classical Information Theory & Processing

• Meaning independent assumption

- R. Hartley, C. E. Shannon

• Probability (Mathematics)-based Information Processing

Shannon’s Communication Theory (1948)

A Communication Model

<table>
<thead>
<tr>
<th>Sender</th>
<th>Receiver</th>
</tr>
</thead>
</table>

\[ x_y \]

Anstraction: (Markov) Stochastic Process

-C. E. Shannon

From Computer Age to Web Age

Man-Machine Interaction

Users

Intensity

Programming (Encoding)

Interpretation (Decoding)

Content

Codes
Form

Computer Network
1. Cognition-based Methods

Cognition-based methods
(Perception-based methods)

refer to a faculty for the processing of information, applying knowledge, and changing preferences.

Symbolic Artificial Intelligence Methods

Production Systems
- A Human-oriented Method
- A Specific Domain Knowledge Representation
- A Common Inference Engine
The Symbolic Domain
Knowledge Representations

Rule-based Representation
(declarative knowledge):

Human disease diagnose
Production Rules:
If \( a, \ldots \) symptoms (fuzzy)
\[ \rightarrow \]
CF: certainty factors
Then \( b \) function disorder (fuzzy)

The Advantages & Disadvantages

- **Advantages:**
  Human-based, Understandable

- **Disadvantages**
  1. Domain Dependent
  2. Deliberative behaviors except common sense,
     natural language understanding,…
     Perception is not applicable

Syntactic Analysis
- The symbolic representation of images

Detector
Semantically meaningful features:
Boundary, shape,..
There is no clear boundary among parts
Segmentation Problem
Descriptor
Structural uncertainty


The Symbolic Representation
- Image Segmentation

If a \_\_ \_b,\_\_\_\_\_ then
o (horse)

Where is the object?
What is the object?

Chicken or Egg?
How does the classical information theory to deal with the semantic aspect of image processing?

2. Statistical Methods
   - Machine-oriented Methods

\[
F(W, D) \xrightarrow{\text{Function Regression}} S
\]

\[
\text{Data Space} \quad \text{Feature Space} \quad \text{Semantics}
\]

Computer representation \hspace{1cm} Human perception
Classical Statistics Theory
-Non-deterministic solution

Law of large numbers in function spaces
Parametric Statistics

Assumption: a known function with a few unknown parameters

\[ ax^2 + bx + c \]

Three Milestones in Machine Learning Theory

Law of large numbers in functional spaces

Learning process

- The necessary and sufficient conditions for uniform convergence
- The sufficient condition for fast rate of convergence
- The necessary and sufficient conditions for fast convergence for any probability measures
\[ F(x, y) = F(y|x)F(x), \quad y = f(x) \]

Data \[\bullet\] Function Semantics

\((\omega_i, x_i), \ldots, (\omega_f, x_f)\) optimization \(F(x, y)\) or \(y = f(x)\)

\[ R(\alpha) = \int L(\omega, \phi(x, \alpha))dF(\omega, x) \]

\[ \phi(x, \alpha), \alpha \in \Lambda, \quad \alpha: \text{a set of scalar quantities, vectors, abstract elements} \]

If \(F(x, y)\) or \(f(x)\) exists

Then, data can infer the function in the probabilistic sense

The Real Problems

[Image of a horse]

Mapping ?

Concepts
Semantics

\(X\)-representation
(Data)

\(S\)-semantics
Mapping

\[ F (W, D) \rightarrow S \]

- Does the mapping (function) between the feature space representation and its meaning exist?
- Only exists in a certain “dataset” for a certain representation (document, image, speech,...)!

A Normalized Alignment Frontal Face Data Base

Extended Yale B
2414 frontal-face
with different lighting
38 individuals
192x168 pixel image
Feature space dimensions:
30, 56, 120, 504

J. Wright, et al. Robust face recognition via sparse representation, IEEE PAMI 09, 31(2):210-227
Anti-Corruption

SRC-Sparse Representation-based Classification, NN-nearest neighbor, NS-nearest subspace, Extended Yale B. (images)

Anti-Occlusion

Percent occluded: 0% 10% 20% 30% 40% 50% 60% 70% 80% 90%
Recognition rate: 100% 100% 100% 100% 100% 90.3% 90.7% 37.3% 7.1%
Low-dimensional Subspace in High-dimensional Space

- The intentionally selective face data base
- The Structure of the sample Space (data space)
- Pixel-based coordinates

The Extended Image Base

Caltech101 (25 categories, 30 images/category)
Bag of (Visual) Words

- Defined in image patches (2005-06)
- Descriptors extracted around interest points (2002-2004)
- Edge contours (2005-06)
- Regions (2005-06)

Detector & Descriptor

Kadir salience region (points)

Histograms of Oriented Gradients (HOG)
-72 dimension

Zuo Yuanyuan, Bo Zhang (2010-)

06.12.2011
Low-level and Local Visual Words (100)

The Sparse Structure in High-dimensional Data Space
Low-dimensional Structure in High-dimensional Space

<table>
<thead>
<tr>
<th>Objects</th>
<th>Precision (SRC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplane</td>
<td>0.947</td>
</tr>
<tr>
<td>Car-side</td>
<td>0.987</td>
</tr>
<tr>
<td>Dalmatian</td>
<td>0.733</td>
</tr>
<tr>
<td>Faces</td>
<td>0.760</td>
</tr>
<tr>
<td>Leopard</td>
<td>0.827</td>
</tr>
<tr>
<td>Pagoda</td>
<td>0.760</td>
</tr>
<tr>
<td>Stop-sign</td>
<td>0.800</td>
</tr>
<tr>
<td>Windsor-chair</td>
<td>0.787</td>
</tr>
</tbody>
</table>

Experimental Results

SVM, 100 visual words

![Graph showing precision vs. object category for different training image sets](image)
**Manifold Learning**

Increase the number of training samples


---

**Scalability**

Complex Landscape !
Generalization Capacity ↓
The Semantic Gap
— The weakness of statistical methods

- The semantic gap between low-level local features and high-level global concepts

  Less semantically meaningful features: colors or their distribution (histogram), gray-values or their distribution, visual words (descriptors from interest points), image patches, image regions, edge, ...

- Lack of structural knowledge

Generalization capacity

Information processing without understanding

The Comparison of Two Frameworks

<table>
<thead>
<tr>
<th></th>
<th>Understandable (Man-machine Interaction)</th>
<th>Knowledge</th>
<th>Uncertainty Management</th>
<th>Scale</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition-based</td>
<td>Good</td>
<td>Y</td>
<td>Poor</td>
<td>Small</td>
<td>Domain dependent</td>
</tr>
<tr>
<td>Probability-based</td>
<td>Poor</td>
<td>N</td>
<td>Good</td>
<td>Large</td>
<td>Data dependent</td>
</tr>
</tbody>
</table>
3. The Combination of Two Frameworks

Statistical Inference

over

An Abstract Structured Knowledge Representation
Contextual Analysis

Markov (language) models - C. Shannon

$n$-gram (item: phonemes, syllables, letters, words, etc. levels)

\[ P(x_i | x_{i-1}, \ldots, x_{i-n}) \]

$n=0$, unigrams (item occurrence frequency)
$n=1$, bigrams
$n=2$, trigrams

2D Flat Structural Models

- Image Region Annotation: horse, sky, mountain, grass, tree

Yuan Jinhui, Bo Zhang (2008-)

Yuan Jinhui, Bo Zhang (2008-)
Region-adaptive Grid Partition

\[ \hat{z}_{\text{argmax}}(z) \]

\[ p(x) \propto \exp \left( -\frac{1}{2} \sum_{i,j} a_{ij} (x_i - x_j)^2 \right) \]

\[ p(z|x) \]

\( I \)-data

\( x_i \)-image position

\( z \)-state (feature)

\( g(z_i, z_j) \)-the probabilistic region-based constraints (co-occurrence)

The Graphical Models

(a) Generative model
(b) Discriminative model
(c) 2D-Hidden Markov model (HMM)
(d) Markov Random Field (MRF)
(e) Conditional Random Field (CRF)
Experimental Setting

- 4002 Corel images (384×256 or 256×384)
- 11 basic (region) concepts
- Features: color moment + wavelet
- 5 models: 2 without structural knowledge (GMM, SVM)
  3 with structural knowledge (HMM*, RMF*, CRF*)

The Categories of Image Region Annotations

<table>
<thead>
<tr>
<th>语义概念</th>
<th>定义及描述</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sky</td>
<td>包含空气，云，烟，雾等</td>
</tr>
<tr>
<td>Water</td>
<td>包括河流，大海，湖泊，喷泉，瀑布等</td>
</tr>
<tr>
<td>Mountian</td>
<td>只含山脉的远景</td>
</tr>
<tr>
<td>Grass</td>
<td>除树木和花朵外的自然植被</td>
</tr>
<tr>
<td>Tree</td>
<td>包含树干，树叶等</td>
</tr>
<tr>
<td>Flower</td>
<td>各种色彩的花朵</td>
</tr>
<tr>
<td>Rock</td>
<td>较近观察的石头，注意与“山脉”的区分</td>
</tr>
<tr>
<td>Earth</td>
<td>自然裸露的地面</td>
</tr>
<tr>
<td>Ground</td>
<td>人加工过的地表，例如道路，广场等，注意与“土壤”区分</td>
</tr>
<tr>
<td>Building</td>
<td>人建造的结构，例如房屋，桥梁等</td>
</tr>
<tr>
<td>Animal</td>
<td>动物皮毛，例如老虎，狮子，大象等</td>
</tr>
</tbody>
</table>
### Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>gmm</th>
<th>svm</th>
<th>humm</th>
<th>nrf</th>
<th>crf</th>
<th>gmm</th>
<th>svm</th>
<th>humm</th>
<th>nrf</th>
<th>crf</th>
<th>gmm</th>
<th>svm</th>
<th>humm</th>
<th>nrf</th>
<th>crf</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sky</strong></td>
<td>.937</td>
<td>.961</td>
<td>.933</td>
<td>.914</td>
<td>.955</td>
<td>.832</td>
<td>.899</td>
<td>.889</td>
<td>.871</td>
<td>.909</td>
<td>.882</td>
<td>.929</td>
<td>.911</td>
<td>.892</td>
<td>.931</td>
</tr>
<tr>
<td><strong>wat</strong></td>
<td>.410</td>
<td>.583</td>
<td>.531</td>
<td>.390</td>
<td>.598</td>
<td>.449</td>
<td>.588</td>
<td>.489</td>
<td>.496</td>
<td>.639</td>
<td>.429</td>
<td>.585</td>
<td>.509</td>
<td>.437</td>
<td>.618</td>
</tr>
<tr>
<td><strong>mmt</strong></td>
<td>.134</td>
<td>.269</td>
<td>.215</td>
<td>.299</td>
<td>.395</td>
<td>.282</td>
<td>.392</td>
<td>.315</td>
<td>.313</td>
<td>.435</td>
<td>.312</td>
<td>.255</td>
<td>.255</td>
<td>.306</td>
<td>.401</td>
</tr>
<tr>
<td><strong>tre</strong></td>
<td>.709</td>
<td>.765</td>
<td>.611</td>
<td>.615</td>
<td>.755</td>
<td>.481</td>
<td>.538</td>
<td>.556</td>
<td>.532</td>
<td>.570</td>
<td>.573</td>
<td>.632</td>
<td>.582</td>
<td>.571</td>
<td>.650</td>
</tr>
<tr>
<td><strong>frs</strong></td>
<td>.475</td>
<td>.591</td>
<td>.584</td>
<td>.561</td>
<td>.615</td>
<td>.513</td>
<td>.694</td>
<td>.447</td>
<td>.411</td>
<td>.695</td>
<td>.494</td>
<td>.639</td>
<td>.507</td>
<td>.474</td>
<td>.653</td>
</tr>
<tr>
<td><strong>rcs</strong></td>
<td>.033</td>
<td>.088</td>
<td>.198</td>
<td>.242</td>
<td>.220</td>
<td>.132</td>
<td>.281</td>
<td>.216</td>
<td>.189</td>
<td>.341</td>
<td>.050</td>
<td>.134</td>
<td>.207</td>
<td>.212</td>
<td>.268</td>
</tr>
<tr>
<td><strong>ert</strong></td>
<td>.230</td>
<td>.386</td>
<td>.337</td>
<td>.184</td>
<td>.294</td>
<td>.397</td>
<td>.497</td>
<td>.328</td>
<td>.372</td>
<td>.539</td>
<td>.391</td>
<td>.434</td>
<td>.332</td>
<td>.246</td>
<td>.445</td>
</tr>
<tr>
<td><strong>grd</strong></td>
<td>.099</td>
<td>.208</td>
<td>.435</td>
<td>.461</td>
<td>.379</td>
<td>.316</td>
<td>.569</td>
<td>.220</td>
<td>.187</td>
<td>.509</td>
<td>.151</td>
<td>.305</td>
<td>.292</td>
<td>.265</td>
<td>.424</td>
</tr>
<tr>
<td><strong>bld</strong></td>
<td>.610</td>
<td>.730</td>
<td>.484</td>
<td>.481</td>
<td>.625</td>
<td>.437</td>
<td>.569</td>
<td>.582</td>
<td>.550</td>
<td>.645</td>
<td>.509</td>
<td>.640</td>
<td>.529</td>
<td>.513</td>
<td>.687</td>
</tr>
<tr>
<td><strong>nll</strong></td>
<td>.096</td>
<td>.297</td>
<td>.295</td>
<td>.312</td>
<td>.480</td>
<td>.294</td>
<td>.573</td>
<td>.285</td>
<td>.204</td>
<td>.540</td>
<td>.144</td>
<td>.392</td>
<td>.290</td>
<td>.247</td>
<td>.508</td>
</tr>
<tr>
<td><strong>avg</strong></td>
<td>.395</td>
<td>.503</td>
<td>.479</td>
<td>.459</td>
<td>.560</td>
<td>.435</td>
<td>.578</td>
<td>.455</td>
<td>.438</td>
<td>.600</td>
<td>.414</td>
<td>.538</td>
<td>.467</td>
<td>.448</td>
<td>.579</td>
</tr>
</tbody>
</table>

### Learning From Human’s Brain

[Diagram of human brain with labeled parts: meninges, corpus callosum, optic tract, lateral geniculate body, primary visual cortex, reticular formation, thalamus, inferior olivary nucleus, inferior colliculus, superior colliculus, basal ganglia, red nucleus, substantia nigra, subthalamic nucleus, globus pallidus, caudate nucleus, putamen, thalamus, hypothalamus, hippocampus, amygdala, cerebellum, brainstem nuclei, optic chiasm, pituitary gland, spinal cord, autonomic nervous system.]
Object Recognition

MIT-CSAIL-TR-2006-028  T. Serre

HMAX-sum + max
### The Theoretical Framework

#### Kernel functions

<table>
<thead>
<tr>
<th>Name</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner product</td>
<td>$K(x,y) = \langle x, y \rangle$</td>
</tr>
<tr>
<td>Normalized inner product</td>
<td>$K(x,y) = \frac{\langle x, y \rangle}{|x||y|}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$K(x,y) = e^{-|x-y|^2}$</td>
</tr>
</tbody>
</table>

#### Pooling functions

<table>
<thead>
<tr>
<th>Name</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>$\psi(a(h)) = \frac{1}{</td>
</tr>
<tr>
<td>$l_1$-norm</td>
<td>$\psi(a(h)) = \sum</td>
</tr>
<tr>
<td>Max</td>
<td>$\psi(a(h)) = \max</td>
</tr>
<tr>
<td>$l_\infty$-norm</td>
<td>$\psi(a(h)) = \max</td>
</tr>
</tbody>
</table>

#### A set of filters

- Classification
- Pooling functions

### Latent Hierarchical Model

#### Web Page Extraction

*Name, Image, Price, Description, etc.*

- **Hierarchy**
  - Computational efficiency
  - Long-range dependency
  - Joint extraction

- **Given Data Record**
Experimental Setting

- Web page extraction
- *Name, Image, Price, Description*
- Models
  - Multi-lager CRFs, Multi-layer M^3N, PoMEN, Partially observed HCRFs

Data set: 37 Template
Training: 185 (5/per template) pages, or 1585 data records
Testing: 370 (10/per template) pages, or 3391 data records

Structural Prediction Learning

<table>
<thead>
<tr>
<th>Learning Rules</th>
<th>Classification</th>
<th>Structural Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal Joint Likelihood Estimation</td>
<td>Naïve Bayesian Network</td>
<td>Hidden Markov Model (1966)^1</td>
</tr>
</tbody>
</table>
The Related Publications


Experimental Results

- **Performances:**
  - $F_1 = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$
  - Avg $F_1$:
    - avg $F_1$ over all attributes
  - Block instance accuracy:
    - % of records whose Name, Image, and Price are correct
Page-Level Evaluation

- Supervision Level 1:
  - Leaf nodes and data record nodes are labeled

- Supervision Level 2:
  - Level 1 + the nodes above data record nodes

Related Publications & Patents


Techniques transferred to Microsoft’s search products:

- Theoretical foundation (JMLR 2009, ICML 2008, ICML 2009a);
- Non-parametric Bayesian (ICML 2011).
- Structure learning of Markov networks (NIPS 2010c, SIGKDD 2010);
- Multi-task learning (NIPS 2010d);
Multi-disciplinary Research at Tsinghua University

Center for Cognitive and Neural Computation
Tsinghua University, Beijing

- Computational Neuroscience
- System Neuroscience
- Intelligent Technology and Systems
- Neural Information and Brain-computer interface
- Learning and Memory
- Cognitive Psychology

Thank You!