Composition Systems

S. Geman, D.F. Potter, Z. Chi

Presented by Haibin Ling
12. 2. 2003

Contents

Introduction

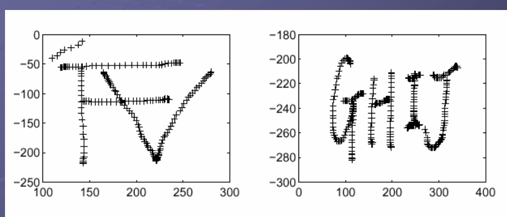
- Definition:
  - Compositionality refers to the evident ability of humans to ***represent entities as hierarchies of parts***, with these parts themselves being meaningful entities, and being reusable in a near-infinite assortment of meaningful combinations.
- Intuitions from the definition
  - Treat objects with tree structures
  - Form objects by composition rules
  - Evaluation candidate objects

What's in this paper

- The purpose of this paper
  - To propose a mathematical formulation of compositionality
    - Tree Structure
    - Composition Rules
  - To devise a probability on the tree structures to promote grouping
    - Inspired by MDL (Minimum Description Length)
    - Recursive grouping
- Framework for recognition
  - Form an object set, contains all candidate objects (trees)
  - Choose the candidate with the minimum code length (MDL)/maximum a posteriori (MAP)

Contents

On-Line Character Recognition

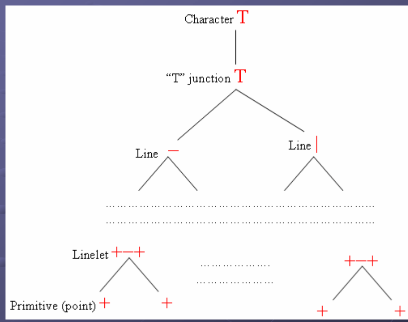


- D. Potter, S.H. Huang, X. Xing

Composition system of uppercase characters

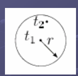
- Object Library Ω :
  - Each object is represented as a tree such that for each non-terminal node *n* with label *l* there exists a composition rule under which the daughters of *n* can bind to form an object of type *l*
  - Ω is the set of all possible objects (trees).
- Primitives *T*: MxM grid points
- Labels: *N*={"A",…, "Z"}∪{ line, linelet, T-junctions, L-junctions, etc.}
  - Linelets: Composed by two primitives
  - Lines: Composed by a line/linelet and a primitive
    Composed by two lines/linelets
  - "T", "L" junctions, etc.
  - Uppercase characters

Tree structure of "T"

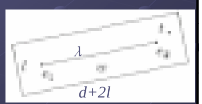


Composition Rules for Linelets and Lines

- Linelets
  - Two points t_1 and t_2 can group to a linelet if their distance is smaller than the given radius threshold *r*



- Lines
  - Composed by a line/linelet λ and a point *t*, e_1 and e_2 are the two points in λ achieve the maximum distance. Point *t* can be grouped with λ if *t* is within the a "bound" rectangle with width *d+2l* and height 2*w*, where d is the distance from e_1 to e_2



  - Composed by two lines

Coding and MDL

- MDL: An optimal interpretation is an assignment that achieves the minimum total description length.
- Example of bits saved of linelet compared to two primitives
  - Code for each type: $\log_2 L$, $L=|N|+|T|$
  - Code for each point: $2\log_2 M$
  - Code for each primitive: $\log_2 L + 2\log_2 M$
  - Code for a linelet: $\log_2 L + 2\log_2 M + 2\log_2(\pi r^2)$
  - Bits saved: $\log_2 L + 2\log_2 M - 2\log_2(\pi r^2) > 0$
    (since $\pi r^2 < M^2$)

On-line Character - Algorithm

- Step 1. Build object library Ω
  - The observed primitives are recursively aggregated into candidate objects under the composition rules.
  - Some sort of pruning based on the description length is used to reduce the size of candidate objects.
- Step 2. Select object from the library Ω
  - Choose a subset of from this collection by choosing successively the next best labeling (minimal description length) among those not been labeled yet, until all the original image is entirely labeled.
  - Use greedy algorithm to make the process fast.

Contents

Labeled Trees Θ

- *T*: primitives (terminals)
  - {0,1,…,255}x{0,1,…,255}
  - Continuous/Discrete
- *N*: labels/types (non-terminals)
- Θ : set of labeled trees

  Each labeled tree is a directed tree graph:
  - Planar, finite and connected
  - Leaf node: labeled with an element in *T*
  - Non-leaf node: labeled with an element in *N*

Composition Rules

- Not all elements in Θ are objects, objects are distinguished by being consistent with composition rules
- Composition rule for label *l* is a pair (B_l, S_l)

  B_l: $\Theta^* \rightarrow R_l$
  - B_l: binding function
  - S_l: binding support, defines allowable values under the composition.
  - Θ^*: set of non-empty strings of labeled trees
  - R_l: **arbitrary** range space

Object Set Ω and Composition System

- The set of objects Ω is the closure of *T* under all composition rules in Θ.
- A composition system is defined as
  - $C = \{ T, N, \{(B_l, S_l)\}_{l \text{ in } N} \}$
- Some properties
  - $T \subseteq \Omega \subseteq \Theta$
  - Any language is attainable from a composition system
  - Can be both context-free and context-sensitive

Example: L-junction, composition rule

$$B_2(\alpha,\beta) = (\theta_\beta - \theta_\alpha, \frac{r_\beta}{r_\alpha}, R_{-\theta_\alpha}\frac{x_\beta - x_\alpha}{r_\alpha}), \qquad L(\alpha) = L(\beta) = 1$$

relative angle

relative length

relative position of endpoints

1 denote label of line

α, β are two lines, r_α, r_β is the length of line α, β, $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

$$S_2 = [3\pi/8, 5\pi/8] \times [.9, 1.1] \times \{(x,y) : \sqrt{x^2 + y^2} < .1\}$$

Example: L-junction, Result

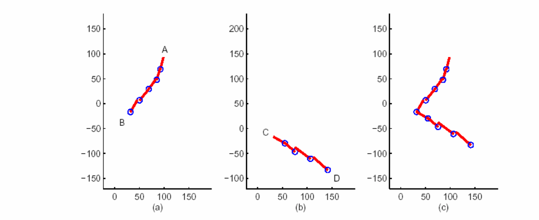


Figure 1.4: L-junction Formation. (a) Line α with endpoints AB, $\theta_\alpha = \angle AB$, $r_\alpha = \|AB\|$, $x_\alpha = B$. (b) Line β with endpoints CD, $\theta_\beta = \angle CD$, $r_\alpha = \|CD\|$, $x_\beta = C$. (c) L-junction $2(\alpha,\beta)$.

Composition System & Context Free Grammar (CFG)

- If the composition rules depend only on labels of the constituents, i.e.

  $B_l(\alpha_1,\dots,\alpha_n) = (L(\alpha_1),\dots,L(\alpha_n))$

  Then it is equivalent to a context-free grammar.
- Attribute function $A: \Omega \rightarrow R$, $|A(\Omega)| < \infty$ and

  $\forall l \in N, \exists A_l : R^* \rightarrow R$

  $w = l(\alpha_1,\dots,\alpha_n) \Rightarrow A(w) = A_l(A(\alpha_1),\dots,A(\alpha_n))$
- If composition rules depend only on attribute values,

  $B_l(\alpha_1,\dots,\alpha_n) = B_l(A(\alpha_1),\dots,A(\alpha_n))$

  Then there is a context-free grammar that produces the same object set as the composition system.

Contents

- Introduction
- Example: On-line Character Recognition
- Composition Systems
- From MDL to Compositional Measures
- Recognition of Scenes in Images

From MDL to MAP

- MDL: *optimal* interpretation = minimum description length
- Two way of encoding
  - Coding an object based on the code of its children:
    - Difficult to identify a sufficient set of attributes that anticipate all possible composition
  - Through probability, start from scratch
    - Recode the constituents in a manner that is natural and particular for the given composition rule.
    - Though **probability measure**
- Distribution P on Ω, favors compositions
  - Shannon code length, $|c(w)| = -\log_2 P(w)$
- Example of coding
  - $w = l(\alpha, \beta)$
  - Bits saved = $|c(w)| - |c(\alpha)| - |c(\beta)| = \log_2(P(w)/(P(\alpha)*P(\beta)))$

Technical Foundation

- From a σ-*algebra* on T (σ_T) to a σ-*algebra* on Θ ($\mathcal{F}$).
  - Skeleton partition
- $\Omega \in \mathcal{F}$ is measurable
- Compositional Measures

Compositional Measures

- A probability measure P on Ω is a compositional measure if

$$P(w) = \begin{cases} Q(w) & w \in T \\ Q(l)Q_l(B_l(\alpha^*))P^*(\alpha^* \mid B_l(\alpha^*)) & w = l(\alpha^*) \end{cases}$$

and $Q_l \ll P_l^*$, for all $l \in N$

where
  - Q is a probability measure on $T \cup N$
  - Q_l is a probability measure on R_l concentrating on S_l, for each $l \in N$
  - $P_l^*(S) = P^*(\alpha^*: B_l(\alpha^*) \in S)$

CFG Case

- When $(B_l(\alpha_1, \ldots, \alpha_n) = (L(\alpha_1), \ldots, L(\alpha_n)))$
- The probability measure is recursively defined as

$$P(\omega) = \begin{cases} Q(\omega) \\ \quad \omega \in T \\ Q(l)Q_l(L(\alpha_1), \ldots L(\alpha_n))P(\alpha_1|L(\alpha_1)) \cdots P(\alpha_n|L(\alpha_n)) \\ \quad \omega = l(\alpha_1, \ldots \alpha_n) \end{cases}$$

$$P(\omega|l) = Q_l(L(\alpha_1), \ldots L(\alpha_n))P(\alpha_1|L(\alpha_1)) \cdots P(\alpha_n|L(\alpha_n))$$

$$Q_l : \{N \cup T\}^* \to [0, 1], \quad \sum_{l^* \in \{N \cup T\}^*} Q_l(l^*) = 1$$

Remarks on Compositional Measure

- Neither the existence nor the uniqueness of a compositional measure is guaranteed.
- $P(\Omega)=1$
- Bits saved:

$$\log_2 \frac{P(w)}{P(\alpha)P(\beta)} = \log_2 \frac{Q(l)Q_l(B_l)}{P \times P(B_l)} \rightarrow \log_2 \frac{Q(l)Q_l(B_l)}{P_l^*(B_l)}$$

  - This can be viewed as the logarithm of a likelihood ratio of two measures on Ω^n
- Recognition: choose the candidates with the maximum probability P
  - Very difficult in calculation
  - Approximation in actual implementations (Huang, Potter)

Example 1

- A Context-Free System
  - T={t}, N={S}
  - $B_S(\alpha^*) = 1$ if α^*=t S→t
    2 if α^*=(S,S) ↔ S→SS
    0 otherwise
  - *Prob*(S→SS) = $Q_S(2) = p$
    Prob(S→t) = $Q_S(1) = q = 1 - p$
  - If $p > 1/2$, there is nonzero probability of producing trees of **infinite** depth, makes $P(\Omega) < 1$
  - If $p <= 1/2$, a unique compositional measure P exists.

Example 2

- Another kind of nonexistence
  - T={t}, N={S}
  - $B_S(\alpha^*) = 1$ for any α^*
- Binding rules are not sufficiently restrictive
  - So $\Omega = \Theta$
  - $P_S^*(\{1\}) = \infty$
  - P exists only when $P(t)=Q(t)=1$

Example 3

- A Context-Sensitive System
  - T={t}, N={S}
  - $B_S(\alpha^*) = 1$ if $\alpha^* = (\alpha, \beta)$, $|\alpha| = |\beta|$
    0 otherwise
  - Ω is the set of balanced trees, the associated language is context-sensitive (Pumping Lemma).
  - A unique compositional measure always exists.

Example 4

- Points, Linelets, and Lines
- Experiments by X. Xing
  - Each rule appended with a formula for computing the gain of encoding
  - Roughly correspond to the negative logarithm (Shannon code length)
  - Actual gains used are more or less ad hoc, not a probabilistic framework yet
  - Brute force search

Contents

- Introduction
- Example: On-line Character Recognition
- Composition Systems
- From MDL to Compositional Measures
- Recognition of Scenes in Images

Recognize Scenes from Images

- Scene: a finite collection of objects.
- Set of scenes

  $$\Psi = \cup_{k=0 \ldots \infty} \Omega_k$$

  $$\Omega_k = \{\{w_1, \ldots, w_k\} : w_i \in \Omega,\ i=1,\ldots,k\}$$
- Recognition problem: find the scene given the image
- Extend a compositional measure, P on Ω to, a distribution D on the set of scenes

  $$D(\sigma) = \Pi_{w \in \sigma} P(\sigma)$$

Discussions

- Advantages:
  - Rich pattern classes (compared to the small number of composition rules).
  - Meaningful parameters
- Problems:
  - Difficult to get accurate estimations of the prior probabilities, Q and Q_l.
  - The need for an automatic inference/parameter estimation procedure.
  - More efficient recognition algorithms is needed.