A Machine Learning Approach to Visual Perception

(Work in Progress)

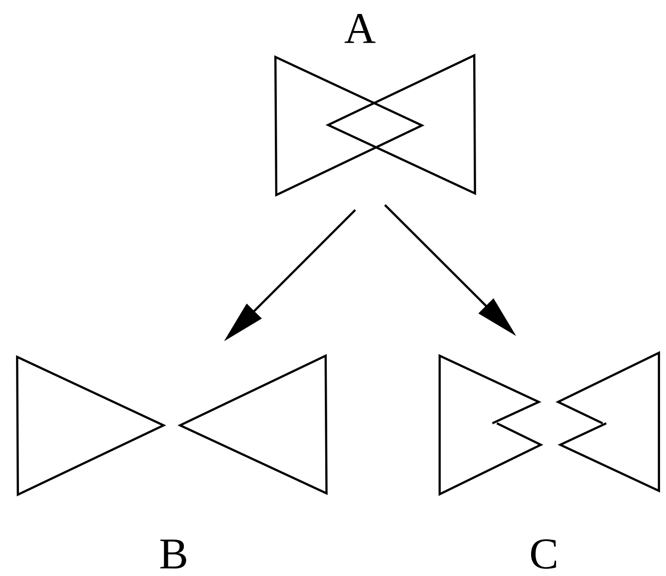
Mehdi Dastani and Khalil Sima'an

Utrecht University

University of Amsterdam

17-19 January 2002

Context and Our Goal



1) ■ ■ ■ ★ ▲ ★

2) ■ ■ ■ ■ ▲ ★

3) ■ ■ ★ ★ ■ ■

Outline

- Structural Information Theory (SIT)
- Likelihood versus Simplicity Principle
- Probabilistic Model for SIT
- Setup of Experiments
- Conclusion & Future Research

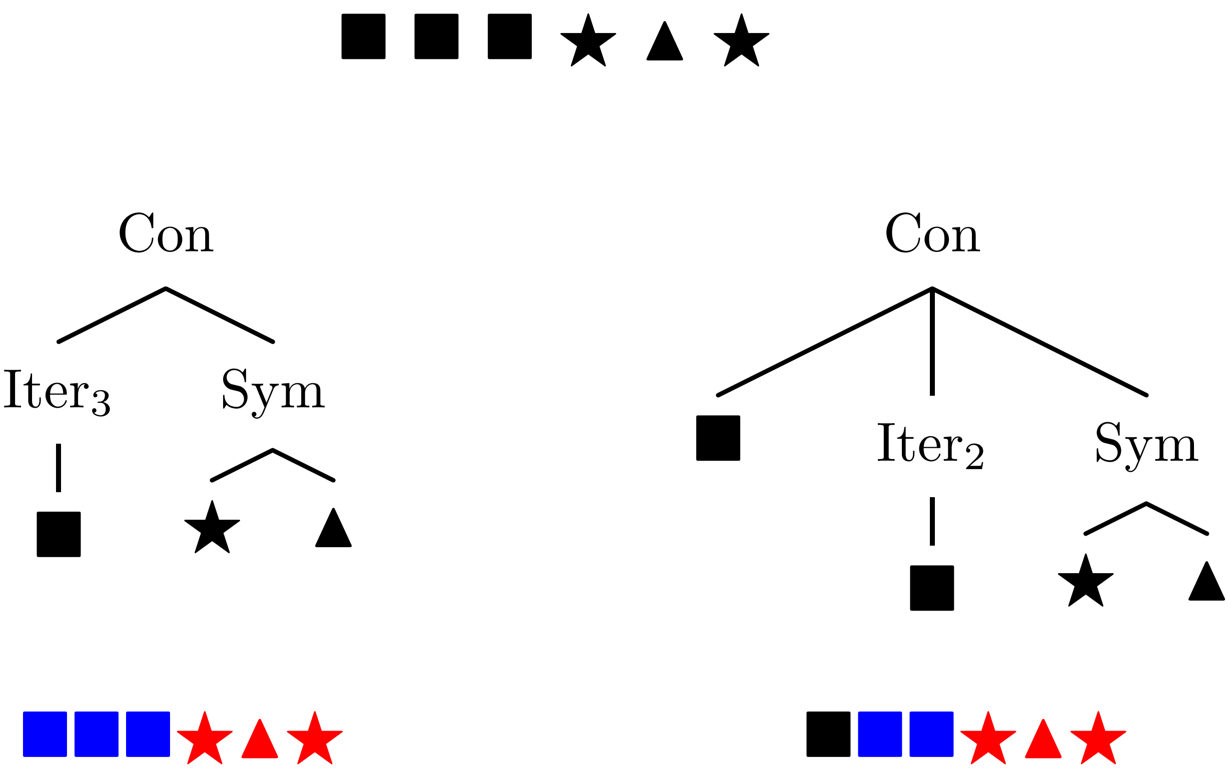
SIT: Structural Information Theory

The class of perceptually motivated regularity are characterized by the ISA operators.

$$
\begin{array}{lcl}
Iter(X\ ,\ n) & \leftarrow & XX\ldots X \\
Sym(X_1\ldots X_n\ ,\ X) & \leftarrow & X_1X_2\ldots X_nXX_n\ldots X_2X_1 \\
Alt_r(X\ ,\ X_1\cdots X_n) & \leftarrow & XX_1XX_2\ldots XX_n \\
Alt_l(X_1\cdots X_n\ ,\ X) & \leftarrow & X_1XX_2X\ldots X_nX \\
Con(X_1\ ,\ \ldots\ ,\ X_n) & \leftarrow & X_1\ldots X_n
\end{array}
$$

SIT: Structural Descriptions

A pattern has different Structural Analysis.



SIT: Information Load

- The Simplicity Principle:

  The Perceived Gestalt of a pattern is reflected by its SIT description which has the Minimum Information Load.

- $IL(A) = \sum$ primitive-elements(A)

- $IL($ Con(Iter(■,3),Sym(★,▲)) $) = 3$

- $IL($ Con(■,Iter(■,2),Sym(★,▲)) $) = 4$

- SIT views visual perception as *compression*.

Inadequacies of SIT

- The information load does not behave correctly for some cases.

★★★■★★★★★★■★★★

★★★■★★★★★★■★★★   ★★★■★★★★★★■★★★

Sym(Con(Iter(★,3),■),Iter(★,3))   Sym(Sym(Iter(★,3),■), $\emptyset$)

Information Load=3   Information Load=2

- Information load is not an empirically defined measure.

Likelihood versus Simplicity Principle

- The likelihood principle:
  The perceived gestalt of a pattern is the one that has the highest probability to occur.
- Likelihood also views visual perception as *compression*.

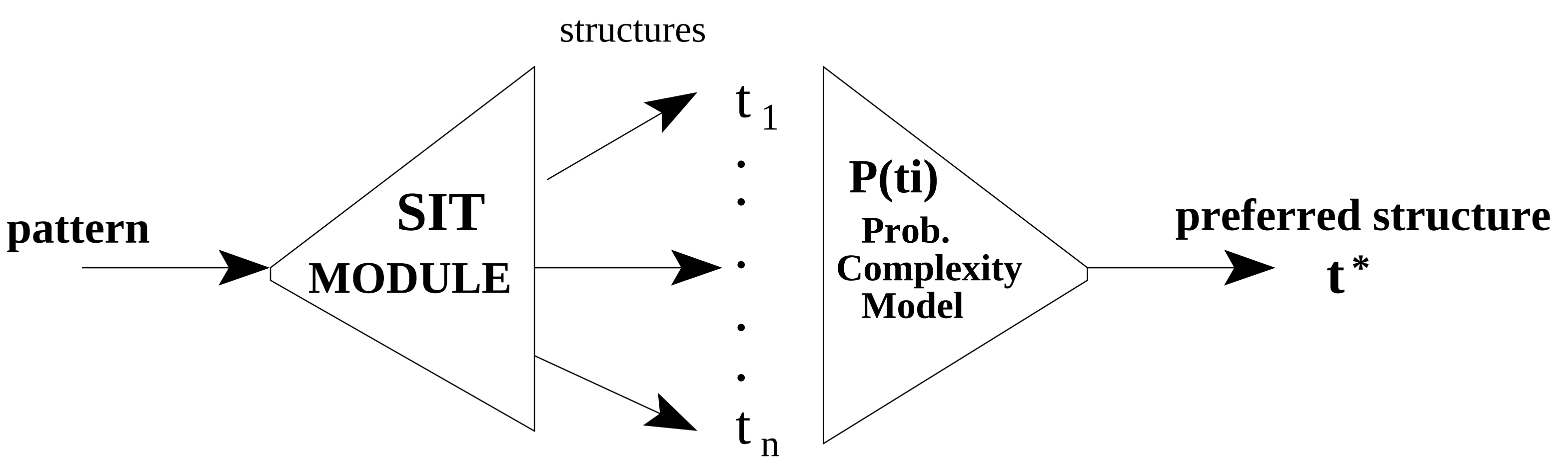
Questions:

- What is the class of gestalts of perceptual patterns?
- What is the probability of these gestalts?
- Enriching SIT with probabilities.

A probabilistic complexity measure for SIT (1)

Given a SIT module over an alphabet Σ.

- Let $SIT(m) = \{t_1, \cdots, t_n\}$ be the set of all SIT structures for a visual pattern m,



- We seek a probability function $P(t_i|m)$.

A probabilistic complexity measure for SIT (2)

- Use $P(t_i)$ for selecting the preferred structure

$$t^* = argmax_{t_i \in SIT} P(t_i|m) = argmax_{t_i \in SIT(m)} P(t_i)$$

- $P(t_i)$ should aim at reducing the expected error over a large sample of patterns.

How to obtain $P(t_i)$?

Statistical learning methodology:

employ a finite sample of $\langle pattern, structure \rangle$ pairs.

How do we obtain this sample ?

How do we approximate $P(t_i)$, for all t_i structures associated with a pattern m from this sample ?

i.e. *since the sample is finite, our model needs to generalize to new, unseen patterns !*

How to obtain a suitable sample D ?

Problems:

- In principle, all visual perception patterns can occur, contrary to e.g. language utterances that are governed by various syntactic and semantic constraints !
- It is hard to collect a sample of visual patterns without introducing unwanted bias.

Hypothesis:

Compression effects in visual perception can be observed in the distributions of the perceived structures rather than in the distributions of the patterns themselves.

Consequence:

The sample D contains a uniform distribution over visual patterns. Probability function $P(t_i)$ will be estimated from the distribution over structures found in D.

Obtaining a complexity measure from sample D.

Probabilistic grammar: extract a probabilistic grammar from finite sample D (analogy with NLP),

How ? two steps:

-Substructures: View every structure t in D as consisting of a sequence of substructures $t = b_0, b_1, \cdots, b_n$,

$$P(t) = P(b_0, \cdots, b_n) = P(b_0) \prod_{i=1}^{n} P(b_i | b_0, \cdots, b_{i-1})$$

-Probabilities: Estimate the probabilities $P(b_i | b_0, \cdots b_{i-1})$ from D.

Analysis of new patterns

- For selecting the preferred analysis of a new pattern m
  - Invoke SIT for obtaining all possible structures of m,
  - Assemble every structure t using substructures b_i from D,
  - Estimate probability $P(t) = \prod_{i=1}^{n} P(b_i | b_0, \cdots, b_{i-1})$.
- What substructures b_i and what context $b_0, \cdots, b_{i-1}$ are suitable for visual perception ?

Simple example

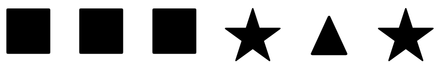
Assumption: A structure consists of a sequence of branches.

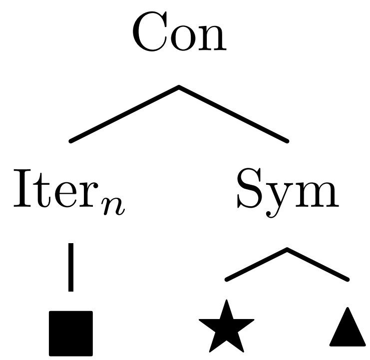
Decompose every structure t in sample D into $b_0, \cdots, b_n$ such that

- every b_i is a branch in t (connects a mother and a daughter node),
- the context of b_i (i.e. $b_0, \cdots, b_{i-1}$) is limited to b_{i-1} (Markovian),
- probability $P(b_i|b_{i-1})$ is estimated through relative frequency in D:

$$\frac{Count_D(b_{i-1}, b_i)}{\sum_{b_j} Count_D(b_{i-1}, b_j)}$$

Example: branch-decomposition





$Con \rightarrow Iter_n$ $\quad$ $Con \rightarrow Sym$

$Iter_n \rightarrow$ ■ $\quad$ $Sym \rightarrow$ ★

$Sym \rightarrow$ ▲

Structurally equal patterns

$\blacksquare\ \blacksquare\ \blacksquare\ \bigstar\ \blacktriangle\ \bigstar \qquad \blacktriangle\ \blacktriangle\ \blacktriangle\ \bigstar\ \blacksquare\ \bigstar$

- Need a function that maps patterns into a canonical form.
- Canonical form, e.g. $A\ A\ A\ B\ C\ B$
- Patterns in sample and new patterns are mapped into this canonical form.

Empirical research questions

- What substructures should be employed ?
- Which local context is most suitable ?
- Is a uniform distribution over patterns in sample D a correct choice ?

Setup of Experiments

- Collect a reasonably large sample of correct structures.
- Train different probabilistic models on sample.
- Evaluate these models and compare to SIT Information load.
- Are there patterns for which humans disagree on perceived structure ?

Conclusion and Future Research

- SIT provides the class of gestalts of perceptual patterns.
- The perceived gestalt of a pattern is the one which has the highest probability.
- Machine learning approach to acquire the probabilities of gestalts.
- Perception versus Cognition: abcghfxyz
- Extending our approach to cover the gestalts of two-dimensional visual patterns.