Toward Communicating Simple Sentences Using Pictorial Representations

Rada Mihalcea, Ben Leong

Department of Computer Science, University of North Texas

{rada,cl0198}@cs.unt.edu

A Picture is Worth 1000 Words

Motivation

Pictures speak a thousand words !

• What do you understand by the following ?

Answer : "The house has four bedrooms and one kitchen"

□ Many languages worldwide (>7000)
- ■some lack formal structures e.g. dialects
- ■capacity to master several languages is limited
- ■long time to learn even one language

□ Globalization
- ■ "breaking down of political, cultural, and trade barriers" (Thomas Friedman)
- ■ require a seamless means of communication

□ Language barriers
- ■ between two speakers of different languages
- ■ children who are preliterate
- ■ people with language disorders
- ■ the deaf/mute

Possible Solution

□ Use pictures !
- ■ body language conveys 90% of information content!
- ■ prehistoric man used non-linguistics or proto-linguistic gestures to make himself understood
- ■ intuitive, minimal learning, potentially universal

Related Work

□ Research on words and pictures :
- **Cognitive Science**
  - word meanings can be understood by tapping on the general-purpose conceptual system that is not restricted to any language
- **Computational Linguistics**
  - use a statistical model to predict word senses from associated images, together with traditional WSD methods
- **Image Processing**
  - combine associated textual semantics with information provided by image features
- **Visual Languages**
  - expression system involving the use of visual objects to express thinking and feeling

Main Resource : Picnet

How it works

- ■Add image representations to concepts defined in WordNet
- ■Encode word/image associations
- ■Combine visual and linguistic representations of world concepts

Typical entry in a dictionary **+ PicNet**

- pipe, tobacco pipe
  - a tube with a small bowl at one end; used for smoking tobacco
- pipe, tobacco pipe
  - a long tube made of metal or plastic that is used to carry water or oil or gas etc
- pipe, tabor pipe
  - a tabular wind instrument

Activities in Picnet

□Administrator functions

□Word/image associations (Web-users)
- ■Free association
- ■Competitive free association (tournament)
- ■Image validation / Scoring
- ■Image donation
- ■Word lookup (search)

Image Validation /Scoring

□ User is shown a synset-image pair – rank its appropriateness.

□Factors to consider:
- ■fitness for the given synset.
- ■quality of the image (size, clarity)

Score based on the user response
- ■ Well related (+4) ; Related to many attributes (+3); Loosely related (+1) ; Not related (-5); Image Upload (+5)

Understanding With Pictures

The Pros

• Universal
• Requires minimal learning
• Intuitive
• Cheap (free contribution by users of PicNet)
• Proven Success (Iconic languages for augmentative communication)

The Cons

• Complex information cannot be conveyed through pictures
  –e.g. "An inhaled form of insulin won federal approval yesterday"
• A large number of concepts with a level of abstraction that prohibits a visual representation
  –e.g. *politics*, *paradigm*, *regenerate*
• Culture differences
  –e.g. some Latin American tribes do not understand the concept of *coffee*

A First Cut

- Simple sentences
  - no complex states or evens (e.g. emotional states, temporal markers, change) or their attributes (adjectives, adverbs)
  - no linguistic structure (e.g. complex noun phrases, prepositional attachments, lexical order, certainty)
  - basic concrete nouns and verbs translated "as is"
- Evaluate the amount of understanding achieved through pictures as opposed to words

Does it Work ?

- Experiments carried out within a translation framework with simple sentences
- A communication process
  - a speaker of an "unknown" language
  - a listener of a "known" language
  - Chinese (unknown) to English (known)
- Three translation scenarios
  1. fully pictorial representations (PicNet)
  2. mixed pictorial/linguistic representations
  3. fully linguistic representations

Sample Pictorial and Linguistic Translations

Evaluation Study

- Interpretations
  - Users asked to provide an *interpretation* based on their first intuition
  - Users' background: Hispanics, Caucasians, Latin Americans
- Data set: 50 short sentences (10-15 words)
  - 30 sentences from language learning courses
  - 20 sentences from various domains (sports, politics,…)
  - Various levels of difficulty
  - 15 (average) interpretations for each sentence
  - One interpretation for each translation scenario
  - Total of 15*3*50=2,250 interpretations

Sample Interpretations

Evaluation

Results

Type of translation	NIST (Bleu)	GTM	Humans
S1 : Pictures	41.21	32.56	3.81
S2 : Pictures + Linguistic	52.97	41.65	4.32
S3 : Linguistic	55.97	44.67	4.40

- Significant amount of information can be conveyed through pictures
  - 76%, compared to the baseline of 0%!
  - Due to the intuitive visual descriptions that can be assigned to some of the concepts in the text
  - Due to humans' ability to contextualize
    - *Read a book* is a more common interpretation than *read about a book*
    - *"He sees the riverbank illuminated by a torch"*

The value of pictures and words …

- S1 (pictures) vs. S0 (no understanding)
  - 3.81 vs. 0
  - role played by images in conveying information
- S2 (pictures with words) vs. S1 (pictures)
  - 4.32 vs. 3.81
  - role played by context that cannot be described with visual representations
  - adjectives, adverbs, prepositions, abstract nouns, verbs cannot be translated into pictures but are important in the communication process
- S3 (words) vs. S2 (pictures with words)
  - 4.40 vs. 4.32
  - advantage of words over pictures in producing accurate interpretations

Conclusions …

□A new paradigm: *translation through pictures*
- □ pictorial translation of basic, concrete nouns and selected verbs in simple sentences

□ Evaluations have demonstrated that the paradigm is:
- ■ universal in generating understanding across different linguistic backgrounds
- ■ comparable to the state-of-art machine translation for linguistic representations
- ■ a significant improvement over the zero baseline of lack-of-communication