Computational Models of Event Type Classification in Context

Alessandra Zarcone[1,2], Alessandro Lenci[1]

⟨a.zarcone@sns.it⟩, ⟨alessandro.lenci@ilc.cnr.it⟩
[1]Università di Pisa, Italy - [2]Scuola Normale Superiore, Laboratorio di Linguistica, Pisa, Italy

LREC 2008 - Marrakech - 28th May / 1st June 2008

A brief outline

Event Types

Event Types

by event type (ET) we refer here to the standard Vendler's classification of predicates (1967)

Event type	[telic]	[durative]	[dynamic]
states	−	+	−
activities	−	+	+
accomplishments	+	+	+
achievements	+	−	+

Table: The features of Vendler's event types

Event Types in Context

1. *John has been reading* **for the whole day** (atelic, durative)
   John has read "The Great Gatsby" **in an hour** (telic, durative)
   John has been reading **papers** *for the whole day* (atelic, durative)
2. *John has been pushing the chart* (atelic, durative)
   John has pushed the chart **to the checkout line** (telic, durative)
3. *The train arrived* **at 5 o'clock** (telic, non durative)
   Europeans *had been arriving sporadically, sometimes with long intervals between arrivals* (atelic, durative)
4. **The path** *goes from the street into the forest* (state)
   The cat *is going to the door* (dynamic)
5. **John** *has hung the picture on the wall* (telic, dynamic)
   The picture *hangs on the wall* (state)

Event Types in Context

Event type polysemy

the ET expressed by a sentence is the result of a **complex interplay** between the verb meaning and its linguistic context

Event type shifts

various contextual factors can shift the verb ET to a new class

Linguistic models of event type hybridism and shifts:

- Dowty 1979
- Bertinetto 1986
- Pustejovsky 1995
- Rothstein 2004

Computational models of event type classification:

- Siegel and McKeown (2000)
- Palmer et al. (2007)

Goal of the paper

task: event type classification in context

hypothesis: the ET expressed by a sentence is the result of a complex interplay between the verb meaning and its linguistic context

aims:
1. model this interplay
2. evaluate the contribution of different types of linguistic indicators to identify the ET of a sentence
3. model various cases of context-driven event type shift

linguistic models: Vendler's event type categories, theories of event type shift

computational models: Maximum Entropy, Self-Organizing Maps

Data

- no corpora annotated for event types available for Italian
- manual annotation of part of the Italian Treebank (Montemagni et al. 2000)
- semi-automatic extraction of contextual features

Verbs	OCCURR.	STA	ACT	ACC	ACH
28	3129	583	430	822	1294

Table: The composition of the corpus for the supervised model

Feature selection

- **adverbial features:**
  - temporal adverbs (“in X time”, “for X time”, etc.);
  - intentional adverbs (“deliberately”, “intentionally”, etc.);
  - frequency adverbs (“rarely”, “often”, etc.);
  - iterative adverbs (“X times”);
- **morphological features:**
  - present tense;
  - imperfect tense;
  - future tense;
  - simple past;
  - perfect tenses;
  - progressive periphrasis;
- **syntactic and argument structure features:**
  - absence of arguments besides the subject;
  - presence of direct object;
  - presence of indirect object;
  - presence of a locative argument;
  - presence of a complement sentence;
  - passive diatesis;
  - subject and direct object, number, animacy and definiteness.

Data and method

- the same lexical item can exibit different ET values in different contexts
- verbs can be more or less polysemous
- groups:

  60%: most frequent ET covers $\leq$ 60% of their tokens
  70%: most frequent ET covers $\leq$ 70% of their tokens
  80%: most frequent ET covers $\leq$ 80% of their tokens
  90%: most frequent ET covers $\leq$ 90% of their tokens

Data and method

Maximum Entropy classifiers (Berger 1996)

$p(a|c)$ is found assuming that the distributions of a set of relevant features $f_i(a,c)$ of c are the only probabilistic constraints involved

task: find the most likely ET given a context c:

$$argmax\left(p\left(a \mid c\right)\right)$$

$$p\left(a \mid c\right) = \frac{1}{Z_c} \prod_{i=1}^{k} {a_i}^{f_i(a,c)}$$

training phase: feature weights estimated with GIS (Generalized Iterative Scaling) algorithm on the training contexts

test phase: weights combined to compute the most likely ET for each new context

Experiment 1

	Baseline	Exp 1
60 % group	56.1%	69.3%
70 % group	60%	72.8%
80 % group	64.6%	75.5%
90 % group	69.6%	78.4%
Whole corpus	79.8%	85.4%

Table: model accuracy

- whole set of features
- most frequent mistakes: non-finite clauses and idiomatic senses

Experiment 2

- to show the contribution to event type classification offered by feature subsets corresponding to specific types of linguistic information
- optimal precision and recall values with the complete feature set

	ACT	STA	ACC	ACH	ACT	STA	ACC	ACH
	adv feat				synt-arg feat			
precision:	0.49	0.35	0.29	0.66	0.89	0.79	0.78	0.86
recall:	0.05	0.1	0	0.14	0.66	0.7	0.92	0.88
f-measure:	0.09	0.15	0	0.24	0.76	0.75	0.84	0.87
	morph feat				whole set			
precision:	0.36	0.38	0.15	0.53	0.84	0.83	0.84	0.88
recall:	0.08	0.62	0	0.49	0.74	0.78	0.89	0.9
f-measure:	0.13	0.47	0	0.51	0.79	0.8	0.86	0.89

Table: Precision and recall results from experiment 2

Experiment 3

- 3 different models, a 2-way classification: *durative* vs. *non-durative*, *dynamic* vs. *non-dynamic*, *telic* vs. *non-telic*
- higher baseline
- baseline outperformed, durativity appears to be the hardest feature to discriminate

	Baseline	Exp 3	Baseline	Exp 3	Baseline	Exp 3
	+/− DUR		+/− DIN		+/− TEL	
60 % group	63.9%	72.8%	60.9%	79.9%	56.9%	79.5%
70 % group	68.3%	74.3%	62.2%	84.9%	66.8%	81.7%
80 % group	75.5%	79.1%	70%	85.4%	71.9%	83.2%
Whole group	88.3%	90.6%	87.7%	92%	84.4%	89.9%

Table: Model accuracy in 2-way classification

Data and method

- 40 Italian verbs from "La Repubblica" (Baroni 2004), high degree of prototipicality with respect to ET
- each verb = distributional vector of co-occurrence frequencies with a number of contextual features
- two verbs with similar context feature distributions = similar ET values

Self-Organizing Maps (SOM, Kohonen 1997)

unsupervised neural network, used to project *n*-dimensional vectors into a 2-dimensional space (map) preserving the topological properties of the input space

Experiment 4

honeycomb map of 100 nodes

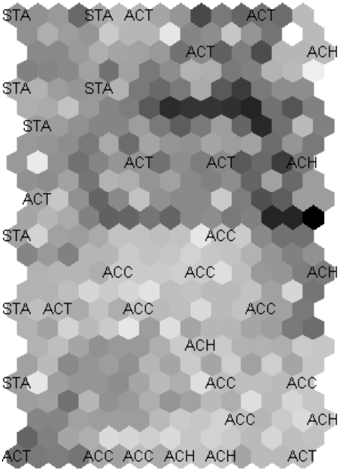

Figure: Experiment 4: the SOM

Experiment 4

honeycomb map of 100 nodes

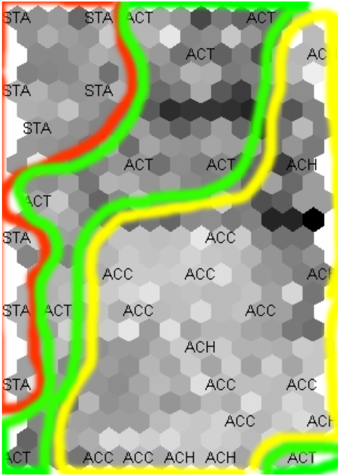

Figure: Experiment 4: the SOM

Experiment 5

- how to model ET shifts with an unsupervised model?
- how to evaluate an unsupervised model for ET classification?

$\Downarrow$

analogy with IR vector space model: an "Aktionsart semantic space", context-driven shifts move a verb towards a new class

training phase: SOM trained as in Experiment 4, then labelled

test phase: SOM used to model context-driven event type shifts on 40 new contexts (for each test item, the nearest labelled BMU is looked for)

Experiment 5

	ACT	STA	ACC	ACH
ACT	1	0	1	0
STA	2	3	4	2
ACC	1	0	6	0
ACH	0	1	9	10
precision:	0.25	0.75	0.3	0.83
recall:	0.5	0.27	0.86	0.5
f-measure:	0.33	0.4	0.44	0.63
accuracy:	50%			

	ACT	STA	TEL
precision:	0.25	0.75	0.78
recall:	0.5	0.27	0.93
f-measure:	0.33	0.4	0.85
accuracy:	72.5%		

Table: Precision and recall results from experiment 5

Conclusions

Event type classification: highly challenging task

- interaction of various contextual factors
- not a trivial task even for humans

Two models: both supervised and unsupervised approaches can account for the contribution of contextual features in identifying the sentence event type

Stochastic algorithms: able to grasp the complex interaction of contextual features (probabilistic cues)

Computational models of ET classification in context can help to

- shed new light on the real structure of event type classes
- gain a better understanding of context-driven semantic shifts