Knowlywood: Mining Activity Knowledge from Hollywood Narratives

Niket Tandon (MPI Informatics, Saarbruecken) Gerard de Melo (IIIS, Tsinghua Univ)

Abir De (IIT Kharagpur)

Gerhard Weikum (MPI Informatics, Saarbruecken)



Legs, person, shoe, mountain, rope..



Legs, person, shoe, mountain, rope..

Rock climbing Going up a mountain/ hill Going up an elevation

Daytime, outdoor activity What happens next?



Legs, person, shoe, mountain, rope..

Rock climbing Going up a mountain/ hill Going up an elevation

Daytime, outdoor activity What happens next? Activity classes Activity groupings Activity hierarchy

Additional information Temporal guidance



Activity commonsense: Related work

ent mining	
cyclopedic KBs: actual e.g. bornOn ntity oriented e.g. Person 1any KBs: e.g. Freebase	
Alexander J. Honnold File File Alexander J. Honnold Her Honnold at the Trento Film Festival (2014) Personal information	
Born August 17, 1985 (age 29) Education UC Berkeley (dropped out)	
Occupation Professional rock climber Climbing career	
Type of • Free solo climber • Bio wall	
Highest grade Redpoint 5.14c (8c+)	
Known for Big Wall Free Soloing Speed record on <i>The Nose</i> of El Capitan	6

Activity commonsense: Related work

Event mining	Commonsense KB	
Encyclopedic KBs: Factual e.g. bornOn Entity oriented e.g. Person Many KBs: e.g. Freebase	Cyc: Manual Limited size No focus on activities	
Alexander J. Honnold	ConceptNet: Crowdsourced Limited size No semantic activity frames	
Personal informationBornAugust 17, 1985 (age 29)EducationUC Berkeley (dropped out)OccupationProfessional rock climberType of climberFree soloClimberBig wallHighest grade Bouldering: V12 (8A+)Known for Ber elever on <i>The Nose</i> of Elever on <i>The Nose</i> of Elever on <i>The Nose</i> of Elever on the solo on the Nose of Elever on the Nose of 	WebChild: No focus on activities	7

Activity commonsense: Related work

Event mining	Commonsense KB	This talk				
<section-header></section-header>	Cyc: Manual Limited size No focus on activities ConceptNet: Crowdsourced Limited size No semantic activity frames WebChild: No focus on activities	<section-header></section-header>				
Known for Big Wall Free Soloing Speed record on <i>The Nose</i> of El Capitan		8				



Activity commonsense is **hard**:

- People hardly express the obvious : implicit and scarce
- Spread across multiple modalities : text, image, videos
- Non-factual : hence noisy



Contain events but not activity knowledge

May contain activities but varying granularity and no visuals. No clear scene boundaries.



Hollywood narratives are easily available and meet the desiderata

EXT. SMALL MOUNTAIN--DAY

Wichita charges up the rockage of a small mountain-hill-type thing. The image repeats itself over and over--each time Wichita is more sweaty, gasping, sneering.

Wichita (V.O.) The rules forbid anyone from the climbing the camp's mountain.













IMS prior W	VN prior	Word match	, VN n score	Selectional restriction score	
maximize $\sum_{i=i}^{j} x_{ij} (\alpha x)$	$\tau_{ij} + \beta$	$B_1 \theta_{ij} +$	$\beta_2 \mathrm{syn}_i$	$\beta_j + \beta_3 \operatorname{sem}_{ij})$	x _{ij} = binary decision var. for word i, mapped to WN sense j
subject to					One VN sense per verb
$\sum_{\substack{j \in S_{\mathcal{V}} \\ x_{ij}}} x_{ij}$ $x_{i_0 j_0}$ $\sum_i x_{ij}$	\leq 1 \leq 2 \leq 1	l r _{ij} l	$\forall i \in V$ $\forall i \in V$ $j mappind i_0 \in X_{ij} \in X_{ij} \in V$ $\forall i \notin V$	$\forall j \in S_{W}, \ ped to \ j' \in S_{V} \ V, j \in S_{V}, \ role-restr(x_{i_0j_0})$	WN, VN sense consistency Selectional restr. constraints
$\stackrel{j}{x_{ij}}$	\in	$\{0, 1\}$			binary decision



Go up an	elevation	Climb up a	mountain	Hike up a	hill	Drink	water
		Participants	climber, rope	Participants	climber		
		Location	camp, forest	Location	sea shore		
		Time	daylight	Time	holiday		

$$\frac{1}{2} \left(\frac{1}{1 + \operatorname{dist}(v_i, v_j)} + \frac{1}{1 + \operatorname{dist}(o_i, o_j)} \right) + \operatorname{Attri}$$

Hypernymy:WordNet hypernymy :+ Attribute hypernymy v_i, v_j and o_i, o_j

Temporal: Generalized Sequence Pattern mining over statistics with gaps #(asynset₁ precedes asynset₂) / #(asynset₁) #(asynset₂)

Probabilistic soft logic - refining Typeof (T), Similar (S) and Prev (P) edges



- 1. Parents often inherit prev. (P) edges from their children: $P(a, b) \wedge T(a, a') \wedge T(b, b') \Rightarrow P(a', b').$
- 2. Similar activities are likely to share parent types $S(a,b) \wedge T(b,b_0) \Rightarrow T(a,b_0)$.
- 3. Likely mutual exclusion between edge types: $T(a,b) \wedge S(a,b) \Rightarrow \neg P(a,b).$
- 4. Siblings are likely to be similar: $T(a,c) \wedge T(b,c) \Rightarrow S(a,b).$
- 5. Similarity is often transitive: $S(a,b) \wedge S(b,c) \Rightarrow S(a,c).$
- 6. Similarity is normally symmetric: $S(a,b) \Rightarrow S(b,a).$



Go up an	elevation	Climb up a	mountain	Hike up a	hill		Drink	wate
		Participating Agent	climber, rope	Participating Agent	climber			
		Location	camp, forest	Location	sea shore	-		
		Time	daylight	Time	holiday			

Tie the activity synsets

Break cycles

Resultant: DAG

Recap

- Defined a new problem of automatic acquisition of semantically refined frames.
- Proposed a joint method that needs no labeled data.



Evaluation

Knowlywood	Statistics
Scenes	1,708,782
Activity synsets	505,788
Accuracy	0.85 ± 0.01
URL	bit.ly/knowlywood

#Scenes is aggregated counts over *Moviescripts, TV serials, Sitcoms, Novels, Kitchen data.*

Evaluation: Manually sampled accuracy over the activity frames.

Evaluation: Baselines

- No direct competitor providing activity frames.

KB Baseline: Our semantic frame (rule based) structure over the crowdsourced commonsense KB **ConceptNet**

Methodology Baseline: A rule based frame detector over our data and other data using an open IE system **ReVerb**

KB Baseline

You open your wallet hasNextSubEvent take out money Normalized domain: concept1 ~ verb [article] noun Organize and canonicalized the relations as follows:

ConceptNet 5's relations	We map it to
IsA, InheritsFrom	type
Causes, ReceivesAction, RelatedTo, CapableOf, UsedFor	agent
HasPrerequisite, HasFirst/LastSubevent, HasSubevent, MotivatedByGoal	prev/next
SimilarTo, Synonym	similarTo
AtLocation, LocationOfAction, LocatedNear	location

Methodology Baseline

Reverb, an openIE tool extracts SVO triples from text

- S and O are only surface forms.
- V is not categorized into a relation. We use a Bayesian classifier to estimate the label of V
- The estimates come from MovieClips.com that provides 30K manually tagged popular movie scenes like, **action**: singing, **prop**: violin, **setting**: theater

$$P(\text{class}|\text{word}) = \frac{P(\text{class},\text{word})}{\sum_{w_i} P(\text{class},w_i)}$$

Methodology Baseline

Reverb, an openIE tool extracts SVO triples from text

- S and O are only surface forms.
- V is not categorized into a relation. We use a Bayesian classifier to estimate the label of V

The estimates come from MovieClips.com that provides 30K manually tagged popular movie scenes like, **action**: singing, **prop**: violin, **setting**: theater

MovieClips tag	Knowlywood attributes	Example
action	activity.v	cut
prop	activity.o	knife
setting	location	bar
occasion	time	thanksgiving
charactertype	participant	policeman



	# activities	
Knowlywood	~1 M	High accuracy & high coverage
ConceptNet based	~ 5 K	High accuracy & low coverage
Reverb based	~ 0.3 M	Low accuracy & high coverage
Reverb clueweb	~ 0.8 M	Low accuracy & high coverage

Visual alignments

~30,000 Images from movies, and additionally, >1 Million images via Flickr tag matching:



External use case -1 : Semantic indexing

Given: participant, location and time Predict: the activity Ground truth: Movieclip's manually specified activity tag.

	MRR	Hit rate	Atleast one hit in Top 10 predictions
ReVerbClue	0.070	0.180	
ConceptNet	0.143	0.345	
ReVerbMCS	0.254	0.415	
Knowlywood	0.327	0.610	

External use case 2: Movie Scene Search

mark go hunting with sophie 's dad.	jeremy go hunt-		Carlos	s an	nd Susa	n are	still	paintin	g over
ing with sophie 's dad. mark tries to	kill a bird . the	the	graffiti	on	the wal	l as	those	people	discuss
man injures it simply. the man tries t	To]	Kill a M	ockin	ng Bird	, hov	vever	, while	talking	
from Sitcom: Peep Show	İ	from	TV serie	s: Desj	perate	Housewiv	ves		

Method: A generative model encoding that a query holistically matches a scene if the participants and activity fit well with the query.

$$P(q|s_t) = \sum_{a \in K} \sum_{p \in A_p} P(q|a) \cdot P(a|p) \cdot P(p|s_t)$$

Conclusion







External use case -1 : Semantic indexing

Given: participant, location and time Predict: the activity Ground truth: <u>Movieclip's</u> manually specified activity tag.

	MRR	Hit rate 🔹	<u>Atleast</u> one hit in Top 10 predictions
ReVerbClue	0.070	0.180	
ConceptNet	0.143	0.345	
ReVerbMCS	0.254	0.415	
Knowlywood	0.327	0.610	

Thank you! Browse at bit.ly/webchild