



Query Expansion in Complex Event Detection with Zero Examples

Maaïke de Boer^{1,2}, Klamer Schutte¹ and Wessel Kraaij²

maaike.deboer@tno.nl, klamer.schutte@tno.nl, w.kraaij@cs.ru.nl

¹ Department of Intelligent Imaging, TNO (Den Haag, The Netherlands)

² Institute for Computing and Information Science, Radboud University Nijmegen (The Netherlands)

Introduction

- Query expansion is a way to bridge the semantic gap between object labels and human semantics of a user query
- Common knowledge bases such as ConceptNet and Wikipedia are easy to access, but noisy
- Expert knowledge bases contain specific information, but requires a lot of (manual) effort
- What method is best to use when in complex event detection?

Query (Event name)

- Match query terms (nouns and verbs) to concept detector labels
- Assign equal weight to matches

ConceptNet

- No matching query term: use ConceptNet to search for the term
- Extract all words with a relation to the term
- Use score of relation in weight

$$W_{CN} = \left(\frac{score_{rel}}{30}\right)^3$$

Wikipedia

- Parse first section of Wikipedia page of the event with Stanford parser
- Weight is adapted from TFIDF to correct for prevalence of concepts

$$W_{wiki} = \frac{f(t, d) \cdot \log \frac{N}{1 + |\{d \in D : t \in d\}|}}{\log \frac{N}{T}}^2$$

- Document set for IDF is an in-domain set of 5798 Wikipedia pages. Terms from the synopsis of the Research set are searched for in Wikipedia and matching pages are included in the set.

Expert

- Manual extraction of nouns and verbs from whole textual description
- Weight is equal to W_{wiki}

Conclusions

- Using Query Expansion not necessarily improves performance, particularly in cases where the whole query or the main noun in the query can be matched to a concept detector (first 11 in Table 1)
- Expert knowledge is preferred over common knowledge
- Expert knowledge is preferred over using only the event query in non-literal events, such as 'wedding shower'
- No significant difference in average performance of ConceptNet and Wikipedia can be found

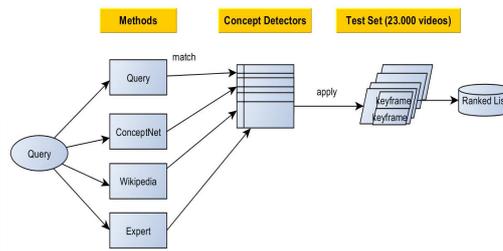


Figure 1. Design

Concept Detectors (VIREO)

- 1000 from ILSVRC-2012
- 346 from TRECVID SIN 2014
- 472 from Research Set

Calculation on Test Set

- Assign highest value over all keyframes of concept detector to video
- Multiply this value with the weight from the methods and sum all contributions

Event_name	Query	ConceptNet	Wikipedia	Expert knowledge
cleaning_appliance	0.12488			0.043775094
rock_climbing	0.14685		0.04525568	0.040022589
fixing_musical_instrument	0.0454			0.01270902
wedding_shower	0.00121	0.055476831	0.06291097	0.154293817
non-motorized_vehicle_repair	0.0136	0.013597247	0.01359725	0.034657089
renovating_home	0.00127	0.001291922		0.001774216
town_hall_meeting	0.00107	0.00123608	0.00105815	0.001361478
winning_race_without_vehicle	0.02594	0.00334611	0.00344042	0.005350132
felling_tree	0.14066	0.028550013	0.06044107	0.010615044
parking_vehicle	0.05384	0.001043524	1.87E-03	0.001766869
tuning_musical_instrument	0.01404	0.015689103	0.01961631	0.010219087
attempting_bike_trick	0.22102	0.149176256		0.252959901
working_metal_craft_project	0.00454	0.002832304		0.0038653
horse_riding_competition	0.07509	0.024312166	0.06756548	0.008353074
playing_fetch	0.0011	0.005295105	3.72E-03	0.002108388
dog_show	0.00496	0.007113812	0.00482721	0.005678951
beekeeping	0.00116	0.001164998	0.01383476	0.006393362
tailgating	0.00133	0.001157935	0.0011567	0.001377089
giving_direction_location	9.50E-04	0.01024546		0.021149057
marriage_proposal	0.00219	0.012782518	0.01403335	0.020424513
MEAN	0.04405	0.019665375	0.02238018	0.031942703

Table 1. Results on Test Set 2014

Further Research

- Temporal Relations:
 - Automatically extract temporal relations (immediately before, before and while) from the methods:
 - Dependency relations of Stanford Parser for Wikipedia and Expert
 - HasSubEvent relation for ConceptNet
 - Take the MAX value of the concept detectors between the keyframes and multiply values
 - While: MAX over same keyframe
 - Immediately before: MAX over two successive keyframes
 - Before: MAX over two keyframes with at least one keyframe in between
- Initial results show slightly improved performance
- Suggestions for improvement?
- Further research in what kind of information improves performance by examining which part of the textual description provides most valuable information