Action Hierarchy Extraction and its Application

Aliaksandr Huminski, Hao Zhang
Institute of High Performance Computing, Nanyang Technological University
Singapore, Singapore
huminskia@iphc.a-star.edu.sg, hao.zhang@ntu.edu.sg

Abstract

Modeling action as an important topic in robotics and human-computer communication assumes by default examining a large set of actions as described by natural language. We offer a procedure for how to extract actions from WordNet. It is based on the analysis of the whole set of verbs and includes 5 steps for implementation. The result is not just a set of extracted actions but a hierarchical structure. In the second part of the article, we describe how an action hierarchy can give an additional benefit in a representation of actions, in particular how it can improve an action representation through semantic roles.

Keywords: action hierarchy, action extraction, semantic role.

1. Introduction

In a natural language an action is mainly described by a verb. Action verbs, also called dynamic verbs in contrast to stative verbs, express actions and play a vital role in an event representation. The key question arises: how to determine if a verb is an action verb? There is a well-known definition that an action verb expresses something that a person, animal or even object can do. Among the examples of action verbs[1], consider the following two: the verb *open* and the verb *kick*.

Meanwhile, this definition creates a mix in understanding. If the verb *open* represents the change of state that happens after some action, the verb *kick* represents the action itself. Rappaport Hovav and Levin (2010) pointed out that an action can be expressed by a verb in 2 different ways. There are verbs called manner verbs that describe carrying out activities – manners of doing: *walk, jog, stab, scrub, sweep, swim, wipe, yell,* etc.; and there are verbs called result verbs that describe results of doing: *break, clean, crush, destroy, shatter,* etc.[2]

It should be underlined that result verbs don't express any concrete action (for example, the verb *clean* doesn't indicate whether it was done by sweeping, washing or sucking; the same way the verb *kill* doesn't indicate how a killing was done) while manner verbs don't express any concrete result (the verb *stab* doesn't define distinctively if a person was injured or killed).

This approach got further elaboration in cognitive science where an event representation is considered to be based on 2-vector structure model: a force vector representing the cause of a change and a result vector representing a change in object properties (Gardenfors, 2017; Gardenfors and Warglien, 2012; Warglien et al., 2012). It is argued that this framework gives a cognitive explanation for manner verbs as force vectors and for result verbs as result vectors. We will further consider ”action verb” as a synonym for ”manner verb”.

The content of this paper is structured as follows. In Section 2 we describe both the general framework for action hierarchy extraction from WordNet and the extraction procedure with the results. Then, in section 3, we describe how an action hierarchy can help in the semantic role representation of actions. Finally, in section 4, we present our main conclusions and the plans for future research in this area.

2. Action Hierarchy Extraction from WordNet

WordNet (WN) as a verb database is widely used in a variety of tasks related to extraction of semantic relations. It consists of verb synsets ordered mainly by troponym-hypernym hierarchical relations (Fellbaum and Miller, 1990). According to the definitions, a hypernym is a verb with a more generalized meaning, while a troponym replaces the hypernym by indicating a manner of doing something. The closer to the bottom of a verb tree, the more specific manners are expressed by troponyms: {communicate}-{talk}-{whisper}.

Meanwhile, troponyms are not always action (manner) verbs although the former is defined through ”manner of doing”. Sometimes they are, like in: {kill}-{drown}. Sometimes they are not, like in: {love}-{romance}.

Action verbs are hidden in the WN verb structure. We know that in some troponym-hypernym relations, the verbs are in fact action verbs. However, there are no explicit ways to extract them yet.

2.1. Framework

Our idea is that action verbs can be extracted from WN if at least one of three conditions, applied to a verb is valid:

1. A verb in WN is an action verb if its gloss contains the following template: ”V + by [...]ing”, where V = hypernym.

2. A verb in WN is an action verb if its gloss contains the following template: ”V + with + [concrete object]”,

[1] http://examples.yourdictionary.com/action-verb-examples.html

[2] Separation of manner and result verbs doesn't mean they fully and exhaustively classify verbs. There are verbs that do not fit in this dichotomy, such as verbs that represent a state, or second-order predicates like *begin* and *start*.

where V = hypernym. Restriction on the concrete object was made to avoid cases like *with success (pleasure, preparation, etc)*.

3. A verb in WN is an action verb if its hypernym is an action verb. In other words, once the verb synset represents action verb(s), all branches located below consist of action verb synsets as well, regardless of their glosses. For example, if {chop, chop up} represents action verbs because of the gloss: *cut with a hacking tool*, its troponym {mince} is also an action verb despite the fact that its gloss doesn't contain any template: *cut into small pieces*.

Let's consider some examples to illustrate conditions 1-3. We start from the top synset {change, alter, modify} (*cause to change; make different; cause a transformation*). It doesn't satisfy the 1st or the 2nd condition, so we go down on 1 level and examine one of its troponyms: {clean, make clean} (*make clean by removing dirt, filth, or unwanted substances from*). It is still not an action verb synset: in the pattern from the 1st condition – "V + by [...]ing" – the verb V (*make clean*) is not a hypernym. On the next level there are synsets with glosses that satisfy either the 1st or the 2nd condition:

- {sweep} (*clean by sweeping*);
- {brush} (*clean with a brush*);
- {steam, steam clean} (*clean by means of steaming*).

So, the verbs *sweep, brush, steam, steam clean* are action verbs. Applying the 3rd condition on them, one can state that all synsets located below these 3 synsets (if any) are action verb synsets. The framework is the basis of the procedure for action extraction.

2.2. Procedure and Results

The procedure[3] includes 5 steps:

1. All verb synsets are automatically extracted from WN 3.1. Total: 13789 verb synsets.

2. At this stage only synsets located on the top level of the hierarchy are automatically extracted. This kind of synsets will be called further "top verb synsets". They have troponyms but don't have any hypernyms. Using this characteristic, all verb synsets extracted on the 1st step have been automatically tested whether they have a hypernym. Total: 564.

3. Top verb synsets are automatically divided into 2 subcategories.

    - The first sub-category is one-level top verb synsets that don't have any other levels below. Examples: {admit} (*give access or entrance to*); {begin} (*begin to speak, understand, read, and write a language*). The reason of extraction is that all 3 conditions mentioned cannot be applied to them. Each condition requires the presence of a hypernym: either to check the patterns (as in the 1st or the 2nd condition) or to define the status of a hyponym (3rd condition). Total: 203.
    - The second sub-category includes all the top synsets left. Total: 361.

4. Top verb synsets from the 2nd sub-category are tested through the conditions 1-3 and the top action verb synsets are extracted. Top action verb synsets are defined as synsets that:

    (a) are satisfied the 1st or the 2nd condition and
    (b) are not satisfied the 3rd condition.

    Top action verb synsets are located on the highest level in action hierarchy.

5. At this stage all the branches from the top action verb synsets are extracted.

The steps of the procedure are illustrated in Figure 1.

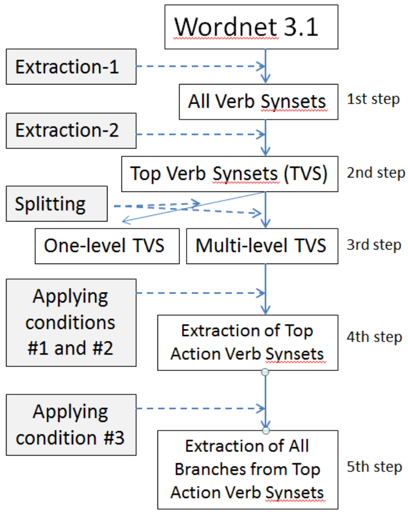

Figure 1: The procedure of action verb synsets extraction.

3. How an Action Hierarchy Can Improve Semantic Role Representation of Actions

As an action is represented by a verb, a semantic representation of actions is closely related to a semantic representation of verbs which has a long history in linguistics. Different approaches and theories consider, as a starting point, either a verb itself, like the theory of semantic roles, or a set of primitives suggested in advance to be combined for a verb representation.

We will further investigate a representation of actions through semantic roles. The aim is to demonstrate how the action hierarchy can help to improve the representation.

As an illustration of the current situation with action representation through roles we take Verbnet (VN) (Kipper Schuler, 2005). It is the largest domain-independent verb lexicon with approximately 6.4k English verbs (version 3.2b). What is important is that all verbs in VN have their role frames. The roles are not so fine-grained

[3] It is a modified procedure of the original one from (Huminski and Zhang, 2018)

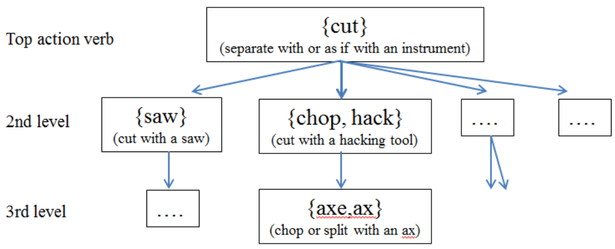

Figure 2: Action verb synsets hierarchy from WordNet.

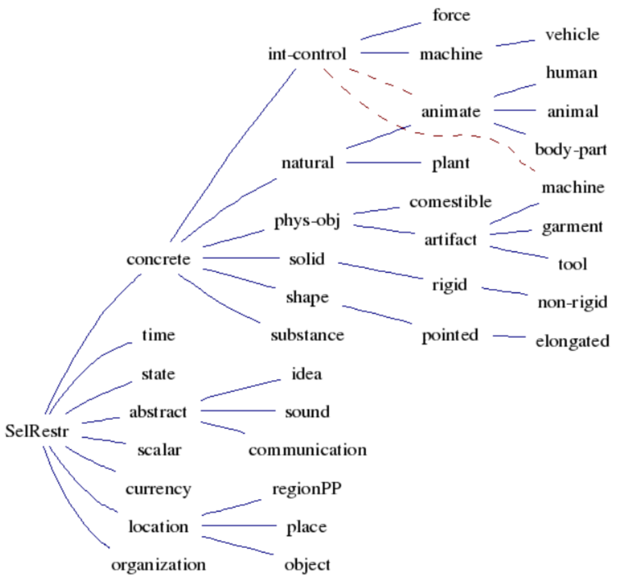

Figure 3: Selectional restrictions in VerbNet.

as in FrameNet (Fillmore et al., 2002) and not so coarse-grained as in Propbank (Palmer et al., 2005). Also VerbNet was considered together with the LIRICS role set for the ISO standard 24617-4 for Semantic Role Annotation (Petukhova and Bunt, 2008; Claire et al., 2011; Bunt and Palmer, 2013).

Let's explore how the action verbs from WN are represented in VN. As an example we take the branch with the top action verb synset {cut}. See Figure 2. In VN the verbs *cut*, *saw*, *chop* and *hack* are located in the class *cut 21.1* (the verbs *ax* and *axe* are not presented) with the other 11 members and the following role frame: {Agent, Patient, Instrument, Source, Result}. This means that 15 verbs of the class are represented the same way and there is no distinction between them. From this point of view an action representation in VN is still coarse-grained. No doubt, it has to be coarse-grained since only 30 roles are used to represent 6.4k verbs.

To make it more articulate, above the roles the system of selectional restrictions is applied in VN. Each role presented in a role frame may optionally be further characterized by certain restrictions, which provide more information about the nature of a role participant. See Figure 3.

For example, the class *eat 39.1* has an agent to be animate and a patient to be comestible and solid. The above-mentioned class *cut 21.1*, to separate it from the other classes, has the following restrictions: {Agent [int_control], Patient[concrete], Instrument [concrete], Source, Result}. Nevertheless, even after applying selectional restrictions, there are classes with both identical role frames and restrictions, without mentioning any distinction between verbs inside a class. For example, the classes *destroy-44* (31 members) and *carve-21.2* (53 members) have the same frame {Agent[int_control], Patient[concrete], Instrument[concrete]}. See Table 1.

<table>
<tr><th>Classes</th><th>Frames</th><th>Members</th></tr>
<tr><td>destroy-44</td><td>Agent[+int_control]+Patient[+concrete]+Instrument[+concrete]</td><td>31</td></tr>
<tr><td>carve-21.2</td><td>Agent[+int_control]+Patient[+concrete]+Instrument[+concrete]</td><td>53</td></tr>
<tr><td>bend-45.2</td><td>Agent[+int_control]+Patient[+solid]+Instrument[+solid]+Result</td><td>23</td></tr>
<tr><td>break-45.1</td><td>Agent[+int_control]+Patient[+solid]+Instrument[+solid]+Result</td><td>24</td></tr>
<tr><td>other_cos-45.4</td><td>Agent[+int_control]+Patient+Instrument+Result</td><td>338</td></tr>
<tr><td>hit-18.1</td><td>Agent[+int_control]+Patient[+concrete]+Instrument[+concrete]+Result</td><td>30</td></tr>
<tr><td>confront-98</td><td>Agent[+animate | +organization]+Theme+Instrument</td><td>18</td></tr>
<tr><td>sustain-55.6</td><td>Agent[+animate | +organization]+Theme+Instrument</td><td>8</td></tr>
<tr><td>begin-55.1</td><td>Agent[+animate | +organization]+Theme+Instrument</td><td>10</td></tr>
<tr><td>stop-55.4</td><td>Agent[+animate | +organization]+Theme+Instrument</td><td>9</td></tr>
<tr><td>establish-55.5-1</td><td>Agent[+animate | +organization]+Theme+Instrument</td><td>25</td></tr>
</table>

Table 1: Verb classes in VerbNet with identical role frames and selectional restrictions.

This may happen because the restrictions are still too coarse for such a big verb data. For example, for the instrument the restriction [tool] located as the final point on the path SelRestr → concrete → phys-obj → artifact → tool is not enough to distinguish the meaning of the 15 verbs from the class *cut 21.1*.

An action hierarchy extracted from WN may benefit the construction of selectional **hierarchical** restrictions (SHR) instead of using just selectional restrictions (SR). Since members of a class in VN are represented in WN in the form of an action hierarchy, we can replace the SR by a fine-grained SHR for each verb in a class. We argue that an action hierarchy will allow improving the semantic role representation of actions by adding more detailed restrictions to a role participant.

Let's consider how an SHR looks like for the class *cut 21.1* with SR [tool] for the role of Instrument. The action hierarchy allows to create SHRs with several levels of restrictions. First, all verbs located below *cut* are under the restriction "instrument for separation". Next step is "hacking tool", "saw", "scissor", "shear", etc. Next one is "whip-saw" (under the "saw"), "ax" (under the "hacking tool"), etc. See Figure 4.

Starting from SR [tool] as a top restriction, an **ontology** of restrictions or SHR is created.

The action hierarchy allows creating a semi-automatic ontology with levels of restrictions, corresponding to the depth of hierarchy in WN.

4. Conclusions and Future Work

In this paper, we offer a procedure on how to extract a hierarchy of actions from WordNet. It can be used for an improvement of the semantic representation of actions.

The procedure of extraction includes 5 steps: 1) extraction of all verb synsets from WN 3.1.; 2) extraction of the top verb synsets; 3) extraction of multi-level top verb synsets; 4) extraction of the top action verb synsets by applying the conditions: "V + by" and "V + with", where V is a hypernym; 5) extraction of all branches of the top action verb synsets using the condition that a verb in WN is an action

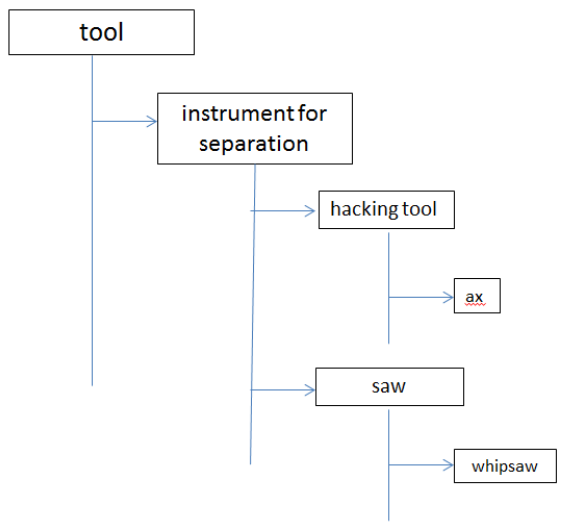

Figure 4: Selectional hierarchical restrictions.

verb if its hypernym is an action verb.
As a result, each branch contains only action verbs in troponym–hypernym relation and thus represents a hierarchy of actions.
Extracted action hierarchy allows improving representation of actions by selectional hierarchical restrictions in a semantic role representation.
As future work, the algorithm can be:

- elaborated by adding new patterns and tuning the original ones. For example, the change-of-state verb synset {die} has a troponym synset {suffocate, stifle, asphyxiate} (*be asphyxiated; die from lack of oxygen*) which clearly indicates the action causing death but the gloss doesn't contain the patterns we are working with.

- enhanced by annotating a set of glosses as to whether they are action verbs or not, to bootstrap machine learning for detecting action verbs from glosses.

5. Bibliographical References

Bunt, H. and Palmer, M. (2013). Conceptual and representational choices in defining an iso standard for semantic role annotation. In *Proceedings of the Ninth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-9)*. Potsdam, Germany.

Claire, B., Corvey, W., Palmer, M., and Bunt, H. (2011). A hierarchical unification of lirics and verbnet semantic roles. In *Proceedings of the 5th IEEE International Conference on Semantic Computing (ICSC 2011)*, Palo Alto, CA, USA.

Fellbaum, C. and Miller, G. (1990). Folk psychology or semantic entailment? a reply to rips and conrad. *The Psychological Review*, 97:565–570.

Fillmore, C. J., Baker, C. F., and Sato, H. (2002). The framenet database and software tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*.

Gardenfors, P. and Warglien, M. (2012). Using conceptual spaces to model actions and events. *Journal of Semantics*, 29(4):487–519.

Gardenfors, P. (2017). *The geometry of meaning: Semantics based on conceptual spaces*. MIT press, Cambridge, Massachusetts.

Huminski, A. and Zhang, H. (2018). Wordnet troponymy and extraction of manner-result relations. In *Proceedings of the 9th Global WordNet Conference (GWC 2018)*, Singapore.

Kipper Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis. Computer and Information Science Dept. University of Pennsylvania. Philadelphia. PA.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Petukhova, V. and Bunt, H. C. (2008). Lirics semantic role annotation: Design and evaluation of a set of data categories. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*. Marrakech, Morocco, 2830.

Rappaport Hovav, M. and Levin, B., (2010). *Reflections on manner/result complementarity*, pages 21–38. Oxford, UK: Oxford University Press.

Warglien, M., Gardenfors, P., and Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics*, 38(3–4):159–193.