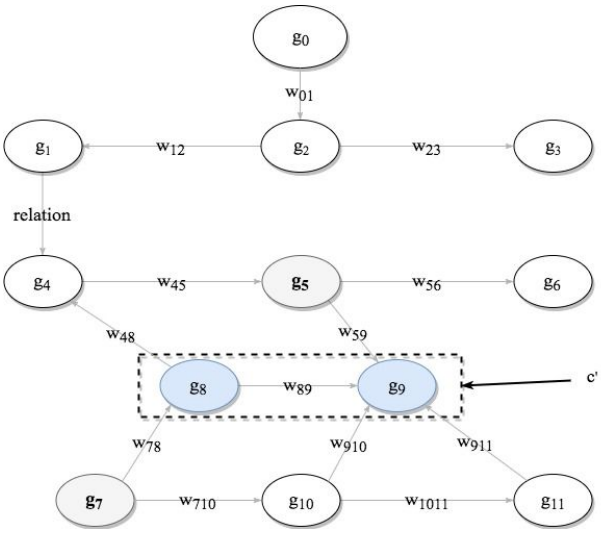
Going Deeper with Semantics: Exploiting Semantic Contextualization for Video Activity Interpretation
Paper ID: 835



In this section, we describe how the contextualization cues are extracted from ConceptNet and the optimal semantic relationships between two grounded concept generators are found.

Formally, let concept generators be represented by g_i for i=1···N and let g_iRg_j represent relations (or assertion) between two concepts. For example, consider the example configuration in Figure 1, which contains a connected structure of generators which represent concepts from the ConceptNet. The generators highlighted in gray i.e. g_5 and g_7 represent the *grounded* concept generators with data evidence from the input video data. The other generators represent the contextual concept generators or *ungrounded* concept generators extracted from ConceptNet. Each of these concept generators are connected through weighted connections or semantic assertions given by w_{ij} for a given generator pair $\{g_i,g_j\}$.

The process of constructing the optimal contextualization cue for two given *grounded* generators g_i and g_j is as follows:

1. Extract the subgraph of all connected concepts from ConceptNet that represent the contextual properties of a given generator g_i.
2. Construct sub-configurations that are representative of all concepts and hence subsequent semantic relationships that are able to connect the two *grounded* generators.
3. Find the optimal sub-configuration that minimizes the energy.

For example, consider the subgraph extracted from ConceptNet for two grounded concept generator g_8 and g_9. The complete sub-graph that comprises of all concept generators related to ***all*** detected *grounded* generators is given by the configuration:

$$c = \sigma\,(g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8, g_9, g_{10}, g_{11})$$

The sub-configuration that represents the contextual information for the generator g_8 is given by the configuration:

$$c_1 = \sigma_1\,(g_0, g_1, g_2, g_3, g_4, g_6, g_9)$$

Similarly the sub-configuration that represents the contextual information for the generator g_8 is given by the configuration:

$$c_2 = \sigma_2\,(g_8, g_{10}, g_{11})$$

The goal is to find the optimal configuration that minimizes the energy of the overall configuration that is representative of the semantic interpretation constructed.

Hence the probability of the sub-configuration that connects the two configurations c_1 and c_2 is given by:

$$P(c'|c_1 \text{ and } c_2) = P(c'|c)$$

Where c' is the sub-configuration that represents the contextualization cues that give the optimal semantic relationships between two grounded generators. The probability of a configuration c is given by the sum of bond energies within the configuration given in Equation 1.

Hence,

$$P(c'|c) = \frac{\sum_{g_i, g_j \in c'} a_{sem}(\beta'(g_i), \beta''(g_j))}{\sum_{g_i, g_j \in c} a_{sem}(\beta'(g_i), \beta''(g_j))}$$

Constructing the optimal contextualization cues for given set of grounded concept generators is a probabilistic induction of the sub-configuration with minimal energy which is reflective of the semantic relationships among the grounded concept generators.