GestureNet: A CommonSense Approach to Physical Activity Similarity

Angela Chang
MIT Media Lab
Authors' address
anjchang@media.mit.edu

Selene Mota
MIT Architecture
Authors' address
atenea@media.mit.edu

Henry Lieberman
MIT Media Lab
Authors' address
lieber@media.mit.edu

Generalizing knowledge about physical movement often requires significant amounts of data capture. Despite the large effort to collect and process activity examples, these systems can still fail to classify movements due to many reasons. Our system, called GestureNet, uses a very small dataset of activity templates to get useful query results for a generalized set of movements. Thus, many more movement profiles can be generated for activity recognition systems and gesture synthesis algorithms.

We demonstrate a system that is able to support a larger set of computer animations based on a small set of base animations. A user can input any motion word recognized by GestureNet, and the system will respond with the closest animation match. GestureNet will also describe the degree to which the new activity is similar to the template profiles. One example is if the user inputs "baseball," the system will show the animation for Run. The commonsense database associates baseball with jogging, which is a type of running. Although the example gesture matrix is small, we demonstrate that our techniques can extend the system to describe variations of these activities (e..g sitting and squatting) which are not currently represented. We can expect that this solution will be useful in application domains where sensor data capture and activity profiles are costly to acquire (e.g. activity classification, animations and visualizations).

Commonsense Reasoning, Animation, Activity Recognition Algorithms.

1. INTRODUCTION

In ubiquitous computing (Tapia, Intille & Larson 2007) and animation studies (Thomas & Johnston 1981), classifying a user's physical movements to a movement profile is an important problem. Once the movement is classified, the system can respond accordingly, e.g. calculate the calories burned while running rather than resting. The approach to robust classification often involves recording large amounts of sensor data for creating profiles of user movements. Collecting large databases of movement profiles is costly and usually this dataset is not appropriate for different contexts (e.g. alternate places or activities).

This paper presents a system that can generalize about activities based on a small number of activity templates (4-6 templates), that abstract properties about the activity such as movement, speed, amplitude and parts of the body involved. We demonstrate that by using only few activity templates, the system can find similarities between new activities and the activity properties specified in the templates by using a commonsense database.

Figure 1. *Some typical activities jump, run, walk and stand*

2. RELATED WORK

Prior efforts in activity classification and recognition have resulted in many methods to capture large amounts of movement data. The set-up and capture of these movements often involves significant time and money. Despite the large effort to collect and process large activity databases, these systems can still fail to classify movements due to lack of enough examples of particular activity templates (Tapia, Intille & Larson 2007).

Conversely, in interface design there has been an effort to recognize specific motions as a way to program by demonstration. By specifying the relationship between sensor input and application

logic, a narrow number of activities are recognized for specific applications (Hartmann *et al.* 2006). The goal of these systems is to quickly and cheaply prototype applications based on a small number of activity templates.

In general, these prior systems cannot recognize more modes of movement outside the envisioned scenarios. We identify an opportunity to unite the constrained data sets with the general relationships provided by commonsense databases, such as ConceptNet (Lieberman *et al.* 2004, Singh 2002).

ConceptNet is an open-source commonsense knowledge base containing over 2.1 million statements of informal conceptual connections. We hypothesize that ConceptNet could be used to augment a sparse activity dataset to allow it to handle activities outside the scope of captured data.

3. THEORY AND METHOD

While ConceptNet contains a broad base of informal knowledge about activities in general, it has little data about how specific physical activities are related about one another. In contrast, our activity data can provide information about the relationships between different activity states. For example, from sensor data, we can note that running is faster than walking and that both activities involve the legs. In order to unite the two information repositories, we use the following method:

First, we start with a dataset of known relationships based on a small amount of activity templates. These relationships are based on general knowledge about the relationships, our experience with sensor data and playing with visualizations of the activity states. (see Figure 1)

Next we create the matrices representing these observations (see Figure 2). The creation was an iterative process by querying ConceptNet and finding matching features that represent the field of knowledge (e.g. activity and movement) that we were interested in. Most of the 2.1 million entries in ConceptNet are not related to physical activity or movement. Thus, ConceptNet queries alone could not provide reliable relationships between our activity states. (For example, walking has a very high number of entries in ConceptNet. Thus, queries about activities referred mainly to walking.)

In order to integrate the general information from our activity dataset with ConceptNet, we had to blend the information from our database with ConceptNet. Unfortunately, simply just blending the two databases did not provide balanced representation from both databases. From ConceptNet, we wanted to deemphasize the large number of irrelevant statements while emphasizing statements about physical activity. Thus, we created custom axes to filter activity-specific information from ConceptNet and, subsequently, we extracted the vectors corresponding to features matching our existing activity templates. Finally, we used the set of extracted axes to compute the similarity between new activities and our activity profiles. See section 4.2 for more detail about this method.

```
def _make_gesture_matrix():
    matrixlist=[]

    matrixlist.append(('run', 'HasProperty','movement'), 50))
    matrixlist.append(('run', 'HasMovement', 'fast'),  50))
    matrixlist.append(('run', 'HasAmplitude','large'), 50))
    matrixlist.append(('run', 'UsesLimbs','legs'), 50))

    matrixlist.append(('walk',  'HasProperty','movement'),10))
    matrixlist.append(('walk', 'HasMovement', 'fast'), 10))
    matrixlist.append(('walk', 'HasAmplitude','large'), 10))
    matrixlist.append(('walk',  'UsesLimbs','legs'), 50))

    matrixlist.append(('stand',  'HasProperty','movement'),-50))
    matrixlist.append(('stand', 'HasMovement', 'fast'),-50))
    matrixlist.append(('stand', 'HasAmplitude','large'),-50))
    matrixlist.append(('stand',  'UsesLimbs','legs'), 50))

    matrixlist.append(('swim', 'HasProperty','movement'),30))
    matrixlist.append(('swim', 'HasMovement', 'fast'), 10))
    matrixlist.append(('swim', 'HasAmplitude','large'), 30))
    matrixlist.append(('swim', 'UsesLimbs','legs'), -50))
```

Figure 2: *The Gesture matrix*

4. APPLICATION

The system has three main components (see Figure 3): the interface, the communication component, and GestureNet. The interface is in charge of demonstrating the activity animations and prompting the user for a new activity query. The communication component is a client-server socket connection that establishes a connection between the interface and GestureNet. GestureNet is an algorithm written that uses ConceptNet and AnalogySpace tools (Speer, Havasi & Lieberman 2008) that computes the similarity between new activities and our profiles.

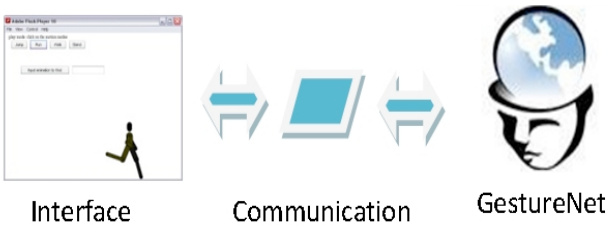

Figure 3: *System design*

4.1 Interface walkthrough

The user begins by clicking four different buttons “Jump, Run, Walk, Stand” to visualize the four activity templates available (see Figure 4). Then, they can enter a new action into a text input box. The query result displays the name of the activity

template that is most like the queried action, along with a degree of similarity among different features. The features demonstrated are movement, speed, uses limbs, and uses arms. The more closely related the feature, the darker the feature appears. Figure 5 shows an example query result for “climb.

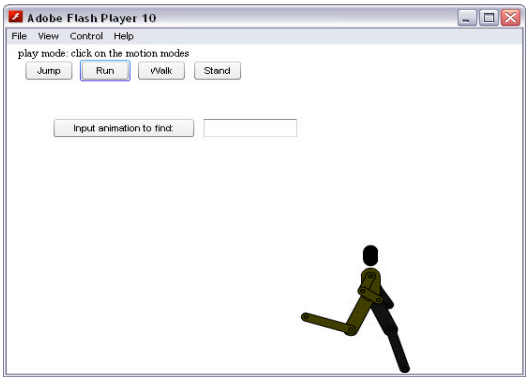

Figure 4: *GestureNet interface at start up. The user can click on the different activity template see the animation.*

```
--------climb--------
similarity cnet vectors:
exercise : 0.595441412603
activity : 0.487983675901
leg : 0.478289691158
physical activity : 0.464099572461
---
similarity over the prototype gesture axis:
jump : 0.291255042006
run : 0.336427150849
walk : 0.371581817853
stand : -0.184311357524
---
similarity new blended vectors:
movement : 0.553877733887
fast : 0.454452098722
large : 0.464114685431
leg : 0.503871703901
arm : -0.485405905235
---
output: walk,0.55,0.45,0.46,0.5,0.0,
```

Figure 2. *The results of entering "climb" shows that it is most like walk, particularly in movement, speed and amplitude of leg movement. Climb is unlike* ***walking*** *in its use of the hands (text for “Uses Hands” is lighter).*

4.2 GestureNet details

GestureNet was built using ConceptNet and Analogy Space tools described by Havasi et al. (2009). ConceptNet is a corpus of commonsense knowledge collected by volunteers on the internet as part of the Open Mind Common Sense (OMCS) project. ConceptNet has more than 700,000 pieces of English-language commonsense data from around 16,000 contributors. The knowledge of ConceptNet becomes computationally useful when it is powered by AnalogySpace. AnalogySpace is a matrix-based representation that uses dimensionality reduction to infer new knowledge about existing relations between a given concept and the information contained in ConceptNet.

In particular, relations between two concepts are described in Concepnet in the form of a 3-tuple (concept, relation, concept). Using this representation, the relationship is broken in two parts: a concept and a feature. The feature is a simple relationship/assertion with one concept. In this way, the information in ConceptNet is expressed as a matrix of concepts and features where positive or negative numbers indicate whether people have made positive or negative assertions about the relationship of a specific concept to a feature, whereas a value of zero indicates that the entry is unknown. Since most entries are unknown, the matrix can be represented sparsely in memory, with the zeroes implied.

If a given concept has unknown value for a feature, but many similar concepts have that feature, then the analogy of the given concept is likely to have that feature as well. AnalogySpace uses these analogies to infer additional pieces of commonsense knowledge that are not part of the original database.

Although ConceptNet contains a large number of entries, it also has a salient amount of noise. For instance, we found that making direct inferences to ConceptNet about physical activity often failed. That is why we used “blending” as an approach to make the reasoning system more robust and compute similarities to integrate the information that was important for our problem.

Specifically, blending is a technique introduced by Havisi (2009) in which two sparse matrices are linearly combined into a single larger matrix that AnalogySpace can analyze using a Singular Value Decomposition method (SVD). When blending is performed, different sources of information can be combined. In our case, we combined ConceptNet with commonsense knowledge obtained by activity abstractions observed in motion studies (see Figure 2).

However, blending alone did not solve the problem of physical activity similarity. We had several problems related to the disparity in size between our activity matrix and ConceptNet. We applied different blending factors to allow both sources of information to contribute to the inference (as in Havasi 2009). However, using the blending factor alone did not work sufficiently well given that entries with large number like “walking” or activities that are also nouns like “stand” were dominating the similarity results. For example, when our custom matrix was emphasized, new activity entries often did not match the right activity profile. In contrast, if we deemphasized our custom matrix and emphasized ConceptNet, many new activities were skewed towards walking or standing which have a large number of entries in the database.
After blending, we realized the need to extract entries relevant to our problem. By filtering relevant entries, we obtain only physical activity relationships from ConceptNet. We created filtering

axes to retrieve entries specific to our activity profiles (see left column of Table 1). This subset of filtered entries are then used to compare a new physical activity to the blend axes used in our existing activity profiles (see middle column of Table 1).

Table 1: *Properties used to extract information in ConceptNet*

Filter Axes	Blend Axes	Activity Templates
IsA, Exercise	Movement	Jump
IsA, Activity	Speed	Run
IsA, Physical	Amplitude	Walk
Activity	Leg	Stand
Leg, UsedFor	Arm	

For example, baseball (Figure 6) and jogging (Figure 7) have profiles that are similar to running, but GestureNet reports that jogging has higher leg activation than baseball, whereas baseball has higher arm activation than jogging.

At the moment, the selection of the activity profile that is most similar to a new activity is made by normalizing the blend axes similarities and computing the differences between each similarity and each activity profile. The chosen profile (from among the four existing templates) is the one with the highest similarity to the new activity. Then we also display the similarity measure across the filtered and blended axes. Figures 6, 7, and 8 show examples of the resulting similarities across filter axes, blended axes, and the activity templates. In the future, we would like to apply more powerful techniques for determining the profile similarity, e.g. using a maximum-likelihood estimation approach MLE (Aldritch 1997).

4. DISCUSSION

Since the commonsense database helps to infer properties of activities never seen before by the system, we consider that these mechanisms can help to fill the gaps in activity recognition and gesture synthesis algorithms. For instance, when asking about new activity, GestureNet can determine the degree to which this new activity is similar to one of the existing profiles in the system. GestureNet can generalize with very few examples by separating out which properties of the new activity are more useful for comparison. Thus, in the case of activity generation, GestureNet can determine how a new activity can be described based on small variations to an existing profile. For example, figure 7 shows the results obtained for "jogging". As can be seen, the system compares "jogging" to the axes contained in our profiles and gives a similarity score. These similarity scores describe relative relationships between the axes (e.g. in jogging legs and arms move, but legs have more activation than arms).


```
--------baseball--------
similarity cnet vectors:
exercise : 0.0999629727147
activity : 0.421200701782
leg : 0.147026110865
physical activity : 0.763550198021
---
similarity over the prototype gesture axis:
jump : 0.046538987793
run : 0.0853891345769
walk : -0.0266952387447
stand : -0.0403682046115
---
similarity new blended vectors:
movement : 0.128775108656
fast : 0.139954717654
large : 0.157032496108
leg : 0.131441943081
arm : 0.226521933431
---
output: run,0.13,0.14,0.16,0.13,0.23,
```

*Figure 3. Baseball is **close** to running. In baseball however, the hands have more activation that in jogging*


```
--------jog--------
similarity cnet vectors:
exercise : 0.791663250623
activity : 0.303176808031
leg : 0.217092785036
physical activity : 0.315357365991
---
similarity over the prototype gesture axis:
jump : 0.241269885552
run : 0.382762392924
walk : 0.336085956662
stand : -0.227078106868
---
similarity new blended vectors:
movement : 0.316235758302
fast : 0.320533255604
large : 0.338456620151
leg : 0.370304493851
hand : 0.228509667059
---
output: run,0.32,0.32,0.34,0.37,0.23,
```

Figure 4. GestureNet query results for jogging

Figure 5. Of the four templates, sit is most closely related to stand due to the lower activation of the similarity blended vectors. Thus, the animation will show the standing, non-moving character when "sit" is entered.

Other interesting examples demonstrate how GestureNet separates out the similarities based on ConceptNet knowledge. For example, the "baseball" query finds that the arm activation is higher than the leg activation. Thus, the closest activity to baseball is running. Another interesting example is that "sit" (Figure 8) is closely related to stand because of the lower activation on all axes.

4. CONCLUSIONS

Our paper demonstrates a method of uniting a very small database containing specific knowledge with a much larger commonsense database. By combining the database using our method, we can describe new concepts according to knowledge about a specific small dataset. Our sample application, GestureNet, can show the relationship of a broad range of activities based on just four activity templates. We demonstrate a novel method of combining blending, filtering axes, and alignment to allow GestureNet to accept flexible input into our system. GestureNet can fill information gaps of missing activity profiles by leveraging ConceptNet to represent new activities based on a small set of existing templates. The closest profile is presented, along with a description of features that are similar between the profile and the queried motion.

Our work is currently restricted to the four activity profiles that we specified in the gesture matrix. However, we demonstrate that our techniques can extend the system to describe variations of these activities (e..g sitting and squatting) which are not currently represented. We can expect that this solution will be useful in application domains where sensor data capture and activity profiles are costly to acquire (e.g. activity classification (Mota & Picard 2003), animations and visualizations (Stiehl, et. Al 2009)).

For future work, we would like to see GestureNet applied to procedural animation, where the degree of similarity for movement features are used to control the animation. Another extension is applying the results to the motion of non-anthropomorphic characters based on feature similarities. We believe that GestureNet's similarity descriptors can be used to aid the classification of physical activities by machine learning algorithms. We can also explore applying GestureNet to different sensor classification environments, such as swimming, object interactions, and household activity recognition (Abowd *et al.* 2002).

3. REFERENCES

Abowd, G. D., Bobick, A. F., Essa, I. A., Mynatt, E. D., & Rogers, W. A. (2002). The aware home: A living laboratory for technologies for successful aging. *Proceedings of the AAAI-02 Workshop "Automation as Caregiver,"* Edmonton, Alberta, Canada, 29 July,1-7, AAAI, Cambridge, MA.

Aldrich, J. (1997). "R.A. Fisher and the making of maximum likelihood 1912–1922". *Statistical Science* **12** (3): 162–176.

ConceptNet Web Site: http://ConceptNet.media.mit.edu/

Divisi Web site: http://divisi.media.mit.edu/

Hartmann, B., Klemmer, S. R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., & Gee, J. Reflective physical prototyping through integrated design, test, and analysis. *User Interface Software and Technology (UIST 2006)*, 15-18 October, 299-308. ACM, New York, NY, USA.

Havasi, C., Speer, R., Pustejovsky, J., & Lieberman, H. (2009). Digital intuition: Applying common sense using dimensionality reduction. Intelligent Systems, IEEE, 24(4), 24-35.

Lieberman, H, Liu,H. Singh,H., Barry, B. (2004). Beating common sense into interactive applications, *AI Magazine, 25*(4): 63-76. AAAI, Cambridge, MA.

Open Mind Common Sense (OMCS) website: http://openmind.media.mit.edu/

Singh, P. (2002) The Public Acquisition of Commonsense Knowledge. *AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, March 2002, Palo Alto, CA: AAAI, Cambridge, MA.

Speer, R., Havasi, C. and Lieberman, H. (2008) AnalogySpace: Reducing the Dimensionality of Commonsense Knowledge, *Association for the Advancement of Artificial Intelligence (AAAI-08)*, Chicago, IL, 13-17 July 2008,548-553, AAAI, Cambridge, MA.

Stiehl, W. Chang, A., Wistort,R., Breazeal, C (2009). The Robotic Preschool of the Future: New Technologies for Learning and Play, *Como 4 Children Competition at Interaction Design for Children (IDC 2009)*, 3-5 June, Como, Italy.

Tapia, E.,M. Intille, S. and Larson, K. (2007) Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor. *International Conference on Wearable Computers (ISWC* 2007), Boston, MA, USA, 11-13 October, 37-40, IEEE. Piscataway, NJ, USA.

Thomas, F., & Johnston, O.(1981). The illusion of life: Disney animation. New York: Hyperion.

Mota, S., & Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. In Computer Vision and Pattern Recognition Workshop, 2003 (CVPRW'03) Conference on (Vol. 5, pp. 49-49). IEEE. Piscataway, NJ, USA.