

# Modeling Information Flow - Humans vs. Computers

## I. INTRODUCTION

We would like to look at QoI in individual nodes in a network and how it is modified by various factors. Using this information, we can more effectively model QoI-aware networks. Then, we also want to explore how these things apply when we look at social networks and/or include humans in the loop of information flow.

## II. QOI MODEL

We want to look at a number of information attributes that all impact the Quality of Information. Specifically, we want to study the following:

- Precision
- Accuracy
- Completeness
- Age
- Timeliness

The first two attributes, precision and accuracy, are determined by inherent qualities of the data with respect to the goal of the application that it supports. One example would be a facial recognition application.

Completeness is a metric that is mostly evaluated within the context of the application's goals and the whole of information being used to support it. For example, if a surveillance application seeks to obtain at least one image of every block in a desired area of a city, then the completeness of a set of images is defined by how many of the blocks are actually visible in the collected images. Further examples are explored in depth in the DCOSS submission.

Age and timeliness are both time-based attributes that each contribute to QoI in different manners. Age describes how long ago the information was generated. The importance of age is highly dependent upon the context in which the information is being used. Timeliness of information describes the time between when it is requested and when it is obtained.

Again, images are good examples for these two metrics. Let's say a known criminal is suspected to have been spotted, but an image of the criminal is needed to compare to the suspected individual. If the application is facial recognition software, then an image that is up to several years old may be acceptable since face structure does not change much, but if a user is attempting to make the identification, then a more recent photo may be necessary to account for changes like change in weight or hairstyle. If we assume that the criminal should be apprehended, then the timeliness in receiving the image with which to identify the suspect is important because it is only useful while the suspect is under surveillance.

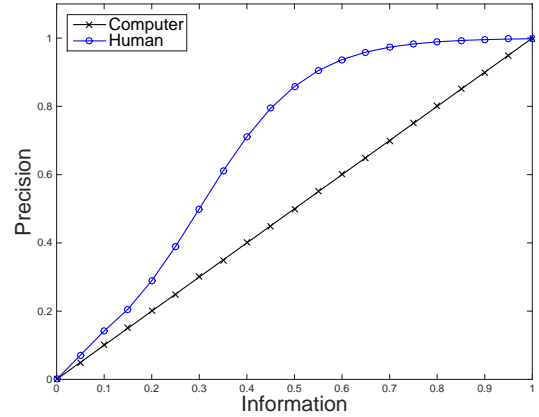


Fig. 1. Precision with more Information in Computers vs. Humans

Now, we want to examine the differences in how each of these QoI attributes is impacted by various effects, such as time, capacity, and load. Furthermore, we want to capture how these impacts occur in computer memories vs. human memories.

### A. Precision

When considering *precision*, perhaps we can look at the following question: How much detail do we have describing the situation of interest? In a scenario with context, humans gain an advantage over computers because of their ability to contextualize new information with previous knowledge and draw conclusions, effectively adding more information. Figure 1 represents this phenomenon showing that as more information is collected, a human is able to provide more detail, increasing the precision above the baseline of the linear accumulation that occurs with a computer.

As time passes, though, humans will forget or merge details, whereas a computer memory is unchanging. Therefore, as shown in Figure 2, the level of detail in human memory degrades over time. Of course, some general information remains, so the precision levels off at some non-zero minimum.

### B. Accuracy

We loosely define *accuracy* as how close a stored data value,  $d$ , is to the corresponding real-world value of  $d'$ , as in [1]. If the value is objective, like the temperature at a location, measuring the accuracy of a value is straightforward.

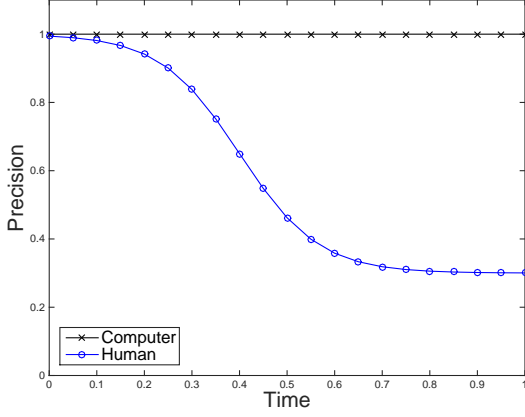


Fig. 2. Precision as Time Passes in Computers vs. Humans

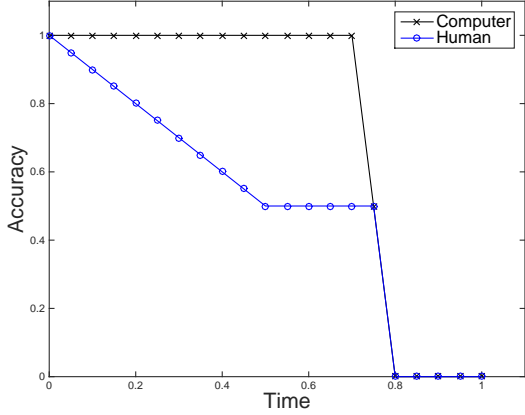


Fig. 3. Loss of Accuracy over Time in Computers vs. Humans

1) *General*: We propose a general model of how accuracy is affected by the time for which things are stored. Since computer memories do not modify any data contents, the accuracy of the data is only affected by its invalidation due to the source creating new data that invalidates the data in the cache. At this point, the accuracy goes to zero. In a human memory, the same timeout occurs eliminating the accuracy of the data at the point in which new information is created, but not yet achieved. The time period before this timeout, however, is different for a human memory because human memories are lossy over time. Therefore, a human begins to lose accuracy of obtained information almost immediately. This loss continues for a period of time, until a threshold of a minimum accuracy that is retained. This minimum accuracy is retained until the information is past the timeout for it to retain any intrinsic accuracy at all. Examples of these two trends are shown in Figure 3.

2) *Facial Recognition*: For most applications that of interest here, like facial recognition for example, accuracy may be best defined as a confidence level or the likelihood of achieving a correct identification with the supplied information.

Using this application as an example, we can compare the

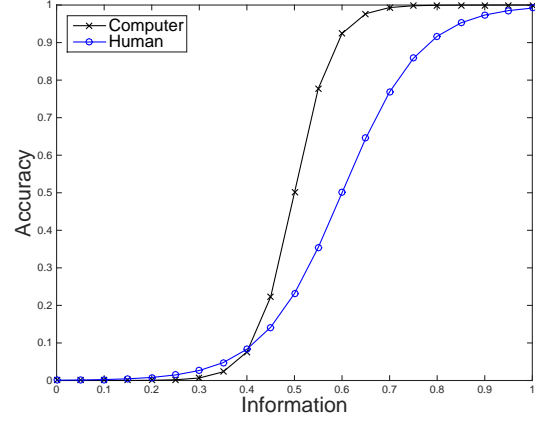


Fig. 4. Accuracy of Facial Recognition with more Information in Computers vs. Humans

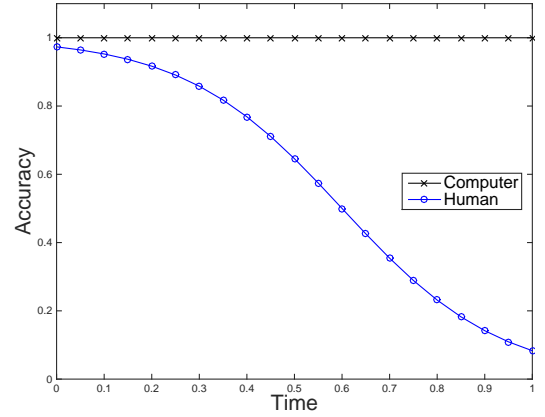


Fig. 5. Loss of Accuracy in Facial Recognition over Time in Computers vs. Humans

effectiveness with increasing information. As shown in [2], image processing techniques can withstand a certain level of compression while remaining almost perfectly accurate. Once a certain compression threshold is reached, though, the ability of the facial recognition application to correctly classify an image drops off dramatically. We propose that the same effect would occur in humans, but that a human's threshold for compression in successful identification is lower, thus requiring them to receive more information, i.e. higher resolution files, to achieve a high likelihood of success. These trends are shown in Figure 4.

### C. Completeness

As in the most recent work (DCOSS 2015 submission), we showed a diminishing returns on completeness as information load is increased. In humans, completeness is quickly gained as information is first received, because they are efficient in pulling important details and obtaining the gist of what is important. Of course, humans have a finite capacity of completeness of information, and experience an overloading after a certain amount of information is received that is not

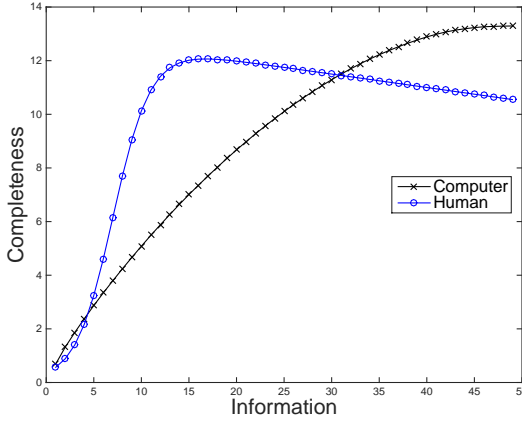


Fig. 6. Information Overload in Computers vs. Humans

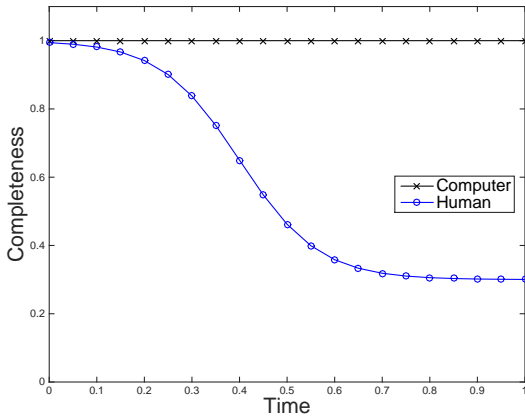


Fig. 7. Loss of Completeness in Human Memory over Time

entirely relevant. These trends are shown in Figure 6.

Figure 7 captures the mutation of completeness of information over time. In a computer, completeness does not change with time, since data is immutable. Again, in a human memory, loss occurs over time, but not completely.

#### D. Age

The age of information simply refers to how long ago the information was generated. This metric would be no different in human memory vs. computer memory.

#### E. Timeliness

Timeliness refers to the delay incurred between a request being generated and the requested information being delivered to the requesting party. Note that this metric is distinct from the age of the information. For example, an image in a database may have been taken five minutes ago or last week. That detail describes the age of the information. If a user connects to the database, and the download time is 30 seconds, then the timeliness of the delivered image is 30 seconds.

We expect that timeliness will be dominated by the network over which information is transmitted, not memory access

time, so we do not outline a difference between human and computer nodes here. (Are there circumstances in which humans require a non-trivial amount of time to recall information that we should consider here?)

### III. NEXT STEPS

We plan to also model the effects of compression on these information QoI attributes. While it is known in computer memories that compression will open up more capacity, the impact of compression combined with the added QoI from more (also possibly compressed) information can be characterized. In human memories, while compression is not a capability in the same way it is in computers, the same effect may arise from humans processing data, extracting only key information. We aim to provide a model of how this occurs and what the resulting QoI is.

Finally, we will incorporate these models of information into networks, designing protocols for caching, routing, and admission control to provide the maximum QoI. Understanding how QoI is impacted by time and load will allow us to study social networks and “human in the loop” scenarios in which information is stored and exchanged through multi-genre networks.

### REFERENCES

- [1] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, 41(3):16, 2009.
- [2] J. Edwards, A. Bahjat, Y. Jiang, T. Cook, and T.F. La Porta. Quality of information-aware mobile applications. *Pervasive and Mobile Computing*, 11:216–228, 2014.