Elaborating Sensor Data using Temporal and Spatial Commonsense Reasoning

Abstract

Ubiquitous computing has established a vision of computation where computers are so deeply integrated into our lives that they become both invisible and everywhere. In order to have computers out of sight and out of mind, they will need a deeper understanding of human life.

LifeNet [1] is a model that functions as a computational model of human life that attempts to anticipate and predict what humans do in the world from a first-person point of view. LifeNet utilizes a general knowledge storage [10] gathered from assertions about the world input by the web community at large. In this work, we extend this general knowledge with sensor data gathered *in vivo*. By adding these sensor-network data to LifeNet, we are enabling a bidirectional learning process: both bottom-up segregation of sensor data and top-down conceptual constraint propagation, thus correcting current metric assumptions in the LifeNet conceptual model by using sensor measurements. Also, in addition to having LifeNet learning general common sense metrics of physical time and space, it will also learn metrics to a specific lab space and chances for recognizing specific individual human activities, and thus be able to make both general and specific spatial/temporal inferences, such as predicting how many people are in a given room and what they might be doing.

This paper discusses the following topics: (1) details of the LifeNet probabilistic human model, (2) a description of the Plug sensor network used in this research, and (3) a description of an experimental design for evaluation of the LifeNet learning method.

1. The Problem Space

The sensor networks of the 1970s and the origins of ubiquitous computing in the late 1980s establish a vision of computation where computers are so deeply integrated into our lives that they become both invisible and everywhere. Realizing this vision requires the building computer systems that exist in our environment and on our bodies; it poses two distinct directions for research: (1) the "human-out," the influence of humanity's needs on technological developments; and (2) the "technology-in," the influence of new technology on humanity. For example, the telephone can be considered as human-out by considering our social need to speak to one another; text messaging on cell phones can be considered as technology-in, since a new technology has affected the way that we express our humanity. Much sensor-network research emphasizes the technology-in direction; the work discussed in this paper attempts to add models of human understanding to sensor networks emphasizing a human-out direction.

LifeNet [1] is a computational model of human life that attempts to anticipate and predict what humans do in the world from a first-person point of view. LifeNet utilizes a general knowledge storage [10] gathered from assertions about the world input by the web community at large. In this work, we extend this general knowledge with sensor data gathered *in vivo*. By adding these sensor-network data to LifeNet, we are enabling a bidirectional learning process: both bottom-up segregation of sensor data and top-down conceptual constraint propagation, thus correcting current metric assumptions in the LifeNet conceptual model by using sensor measurements. Also, in addition to having LifeNet learning general common-sense metrics of physical time and space, it will also learn metrics to a specific lab space and chances for recognizing specific individual human activities, and thus be able to make both general and specific spatial/temporal inferences, such as predicting how many people are in a given room and what they might be doing.

Three topics are covered in the following sections of this paper: (1) details of the LifeNet probabilistic human model, (2) a description of the Plug sensor network used in this research, and (3) a description of an experimental design for evaluation of the LifeNet learning method.

2. LifeNet: a first-person model

LifeNet is a first-person common-sense inference model, which consists of a graph with nodes of common-sense human language phrases gathered from OpenMind Common Sense [10], ConceptNet [2], Korean ConceptNet [**?**], Japanese ConceptNet [**?**], Brazilian ConceptNet [**?**], PlaceLab data [4], and Honda's indoor common sense data [5]. Examples of common-sense knowledge from OpenMind include: "washing your hair produces clean hair"; "shampoo is for washing your hair"; "you can find shampoo in a shower"; etc. This knowledge is related in three ways: logical relationships, temporal probabilistic distributions, and spatial probabilistic distributions. LifeNet might infer that

"I am washing my hair" *before* "My hair is clean."

All of the reasoning in LifeNet is currently based on probabilistic propositional logic; the benefits of this design include: (1) probability eliminates the need to debug very large databases containing millions of strict logical relationships; and (2) higher order logical representations can often be compiled into a propositional form before inference routines are performed, so this compilation feature could be an extension to LifeNet but our application does not require the complexity of first-order logic.

2.1. The LifeNet network architecture

The inference routine that is used in LifeNet is Pearl's loopy belief propagation. This algorithm is used for a number of properties: (1) scalable, (2) distributable, (3) other equivalence-class algorithms exist. The belief propagation algorithm is scalable in the way that the algorithm functions at a fine granularity with respect to data it has to process. Belief propagation runs in roughly linear, $O(n)$, time with the number of nodes, which is important when dealing with the millions of nodes in LifeNet; also, the memory required to implement the belief propagation algorithm is constant, $O(1)$, per node which allows for the second property. The locality of these finely granular data structures for each efficient calculation makes the belief propagation algorithm scalable and distributable in a heterogeneous network of many different processor types and capabilities, which applies to flat as well as heterogeneous sensor networks. Base stations may have a server class processor available with gigabytes of RAM, thus they are able to process millions of nodes, while other processors may be sleeping most of the time and are only able to process on the order of 10 or 20 nodes when they are awake, which would mainly be used for limiting radio communication data between nodes. The third property of equivalence for the belief propagation algorithm refers to the fact that it belongs to a more general class of equivalent algorithms, namely Distributed Hill-Climbing algorithms. This class of algorithms includes the max-product algorithm, recurrent neural networks (or recursive sigmoidal regression networks), distributed genetic algorithms, and others. LifeNet in its present form has been designed partly as a development platform for this class of algorithm, all of which could span many processing nodes of different capabilities spanning decentralized servers to sensor network leaves in the same process.

2.2. Distributed-processing characteristics

Part of learning from everyday experience is our ability to categorize and segregate our knowledge into efficient domains of context-specific ways to think. We have briefly looked into ways to automatically segregate a large LifeNet into multiple domains of context specific ways to think that can be processed independently, allowing for many independent reasoning algorithms to be run in separate processes that communicate a minimal amount of information. A hierarchical partitioning was calculated by iteratively applying spectral partitioning by Chaco **??**.

We are experimenting with graph-partitioning algorithms on the entire LifeNet graph in order to separate very dense inference processing areas of the graph into separate processing modes. Using these techniques to divide the LifeNet processing and communication load across a heterogeneous sensor network has not been attempted, but the belief propagation algorithm has been shown implemented in a sensor network of this sort **??**. Exact inference algorithms in sensor networks such as the Junction Tree algorithm [8] will not scale to large belief networks such as LifeNet.

2.3. Logical truth inference

2 The LifeNet logical inference is based on a collection of truth relationships between statements about a typical person's life. The inference is used by providing LifeNet with evidence in the form of language statements associated with truth values that specify the probability of that statement. The logical model is specified as a Markov random field, which performs roughly the same purpose as the first version of LifeNet [1], except that the model in use now specifies explicit distances between time events rather than simply using a sliced model of time. The details of the temporal inference will be discussed with spatial inference after reasoning about logical truth.

2.3.1. Reasoning with logical evidence

Each probabilistic relationship between LifeNet phrases exists as a tabular probability distribution, forming a propositional Markov Random Field [3]. These relationships relate the nodes within the Markov field. We will refer to these cliques as ψ_i for $i = \{1, 2, \ldots, C\}$ when C is the number of cliques within LifeNet. ψ_i is defined in terms of the probability distribution of the set of variables within that clique, ψ_{iX}. LifeNet factors, ψ_i, are tabular functions of the states of those factors, ψ_{iX}.

A sample tabular potential function for a three-node potential function is shown in figure **??**. The potential functions, ψ, are indexed by the probabilities of their nodes, so although what is stored in each tabular potential are the probabilities of each node being 0 or 1 (*false* or *true*), these potential functions are actually linearly interpolated functions of the probabilities of these nodes, which can take on any values from 0 to 1. These potential functions are calculated as a sum weighted by the probabilities of all possible interpretations of a potential function:

Potential functions can be simplified relative to one variable, X_j, attaining a specific truth value, ν, which if the potential function is a probability table is effectively conditioning on that variable attaining that value. To calculate this potential conditioning, we sum over all possible combination of truth values within the potential function, ψ_i, that contain the condition $X_j = \nu$:

$$\psi_i(X = \nu) = \sum_{\lambda \in \Lambda_i^* \setminus (X \neq \nu)} \psi_i(\lambda),$$

where Λ_i^* is the set of all combinations of binary truth values for the set of variables, Λ_i, of the potential function ψ_i.

For each potential function $\psi_i(\lambda)$, the domain, λ is not a binary space but is instead a bounded real space such that

$$\lambda \in [0 - 1]^{|\psi_i|},$$

where $|\psi_i|$ is the dimensionality of the clique, ψ_i. This function is calculated by making a weighted sum of every tabular entry in the potential. The linear weighting is equal to the probability of that entry being true, given the domain, λ.

$$\lambda(\mu) = \prod_{X \in \Psi} P(X = \mu_X)$$

$$\psi(\lambda) = \sum_{\mu \in \Lambda^*} \lambda(\mu) \cdot \psi(\mu),$$

where λ_i is a set of probabilities for all nodes within the potential function. Potential functions need not sum to one and in general will not because they are not probabilities, but are factors that when multiplied together result in probability distributions.

LifeNet's belief-propagation algorithm accepts evidence, E, for the probability of a subset of the LifeNet nodes. Given this evidence, belief propagation can efficiently estimate the probabilities of the remaining nodes. Let ξ_X^0 be the initial estimate of $P(X|E)$, which is the initial state of the iterating belief propagation algorithm. Within LifeNet, we assume $\xi_X^0 = 0.5$ for all nodes, X, such that $X \notin E$. Our purpose for using the belief propagation algorithm is that it is an efficient albeit unreliable method of iteratively calculating the following limit:

$$\lim_{k \to \infty} \xi_X^k = P(X|E)$$

An efficient unreliable method is used in order to allow us to make the problem of probabilistically reasoning over millions of relationships tractable. For each node, X, we find a new estimate of $P(X|E)$, based on the current probability estimates, ξ_X^k, which gives us ξ_X^{k+1}. At each iteration, the probabilities for the nodes within the *Markov blanket* for each node is assumed to be equal to the most recent probability estimates for those nodes in the blanket.

The *Markov blanket*[1] for a node, X, in LifeNet, or any M.R.F., is equal to the set of cliques that contain that node. The subset of all cliques, ψ, that contain a node, X, is the Markov blanket, X_β, of that node:

$$X_\beta = \{\psi : X \in \psi \in \Psi\}.$$

The Markov blanket of X is the minimal set of nodes that when known, effectively make $P(X)$ independent from any other evidence within the network.

The belief propagation algorithm uses the potential functions by setting the domain of the potential functions at iteration, k, to be

$$\lambda^k(\mu) = P(\mu|E, \xi^k).$$

The iterative algorithm for updating the probability estimates, ξ_i, for each of the nodes is

$$\xi_X^{k+1} = \prod_{\psi \in X_\beta} \psi(\lambda^k), \text{ for all } X.$$

Unfortunately, although the belief propagation algorithm is efficient ($O(n)$ time in the number of nodes), belief propagation is (1) not guaranteed to find the correct distribution when it converges, and (2) not guaranteed to converge. But for LifeNet's current level of complexity the basic belief propagation algorithm converges reliably.

2.4. Inferring past, present, and future.

The temporal and spatial reasoning within LifeNet are now handled as part of a non-parametric belief-propagation algorithm that uses mixtures of Gaussians to represent distributions in real-number spaces. This will be the technology that allows us to incorporate the 9-dimensional sensor space of the Plug network with LifeNet's common-sense spatial and temporal inference.

The spatial representation uses a three-dimensional Gaussian subspace to represent spatial relationships between propositions that can be true or false relative to a position in latitude, longitude, and altitude dimensions measured in meters. For example, if the system were given a list of objects that are known to exist together, LifeNet can provide a probability distribution over all possible arrangements of those objects.

[1] *Markov blanket*: In this paper, we refer to the Markov blanket at the set of cliques that a given node belongs to because this is easier within the M.R.F. framework, but in general the Markov blanket is referred to as the set of nodes that are contained within these cliques. Or more generally, the set of nodes when whose probabilities are known fully specify the probability of a given node such that

$$P(X|X_\beta) = P(X|X_\beta, E)$$

for any evidence, E.

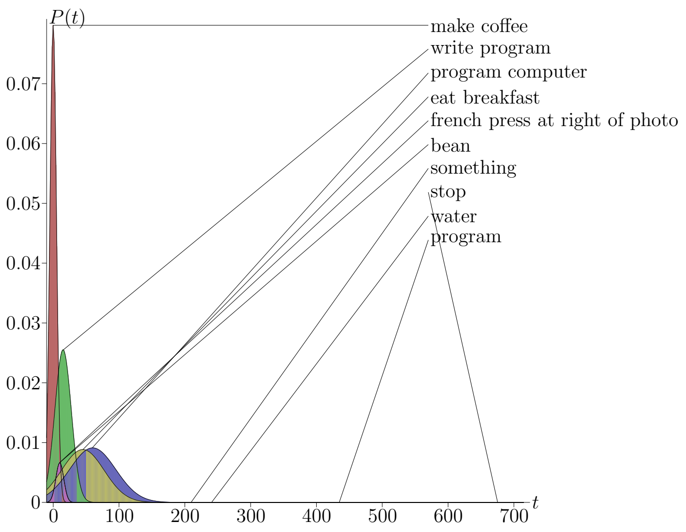

Figure 1: The “make coffee” node is set to have a probability distribution in time that is a simple Gaussian with $\mu = 0$ minutes and $\sigma = 5$ minutes. Time, t, is in minutes. The subsequent distributions were generated by LifeNet’s common sense knowledge and assumptions (to be corrected by sensor data) of temporal distance.

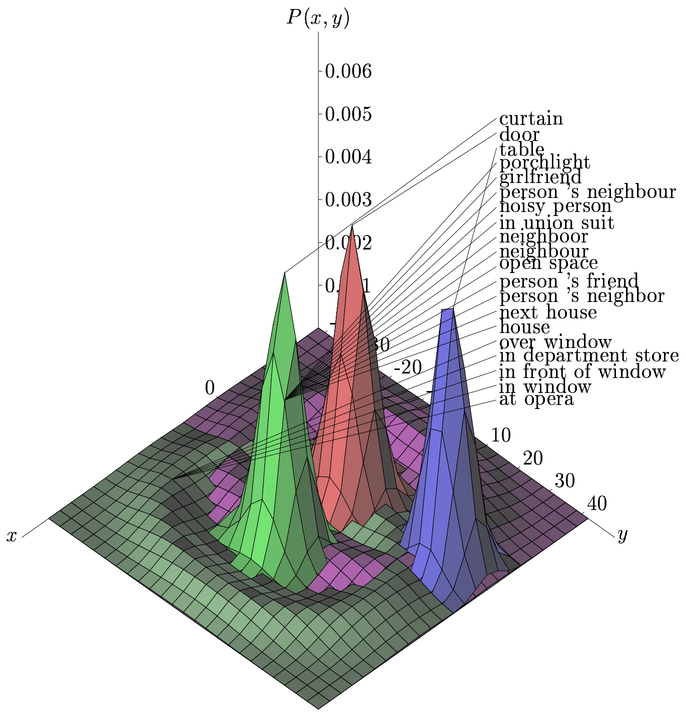

Figure 2: The “window”, “curtain”, and “door” nodes are set to have probability distributions that are simple Gaussians in two-dimensional floor space measured in meters.

The inferred probability distributions for the other concept nodes are shown as mixtures of Gaussians that are in this case circular, but can be in general an approximation of any distribution.

3. The Plug sensor network

The sensor network that we are using for both learning common sense and for recognizing and predicting human behavior is the Plug sensor network [6]. This network is a heterogeneous network consisting of base-station powerstrips that contain 9 sensor modalities: sound, vibration, brightness, current (through 4 separate plugs), and wall voltage. the Plug sensor network will be augmented by small low-power nodes with accelerometers that can be used to roughly position and track individual common objects around our lab space, which has the base station plugs scattered throughout. Using this sensor network to monitor how individuals interact with their physical environment by moving specific objects or simply by their sensor impression on the environment provides a stream of data that can be correlated with simple conceptual human language descriptions of the same events so as to define a supervised probabilistic learning problem. the Plug sensor network is a useful device that theoretically could be readily deployed in both home and office settings.

4. Common Sense Activity Recognition

Because LifeNet has already incorporated millions of semantic relationships from other common sense knowledge databases, the existing context that this semantic knowledge will provide in the learned relationships between sensor events will be the novel aspect of our approach to the problem of sensor network event recognition. Incorporating sensory data into LifeNet’s common sense knowledge will provide a rich source of temporal event sequences and concurrencies, which will also add the specific distance relationships between those data. LifeNet will use what limited context it can infer from the raw sensor data in order to provide more context for further sensing the environment. This technique of using context to narrow the scope of the sensor network could focus the battery energy of the sensor network on specific sensor modalities at a certain times that would be important for a type of resource limited top down inference to take place.

Let us consider a jogger that wants to use a device that can be carried or otherwise worn in order to remember a common sense description in human language of what is going on around her. Simple sensors do not tell us this information directly. Simple sensors on the human body can detect a number of dimensions of data in time, such as temperature, light level, sounds, vibrations, accelerations, and also electrical measurements (EKG, EEG, BCI). So, when she wants to see at the end of the day when she was “jogging” the system can respond with when the sensor data most likely reflects “jogging” as it is related to other common sense concepts, which are in turn related to raw sensor data.

Acknowledgments

This work is supported in part by a grant from *France Télécom R&D*.

References

[1] Singh, P.; Williams, W. “LifeNet: a propositional model of ordinary human activity” Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture; DC-KCAP, 2003

[2] Liu, H., Singh, P. “ConceptNet: A Practical Commonsense Reasoning Toolkit” B.T. Technology Journal; Volume 22, forthcoming

[3] Weiss, Y.; Freeman, W. “Correctness of belief propagation in Gaussian graphical models of arbitrary topology” University of Berkeley California; Report No. UCB/CSD-99-1046, 1999

[4] E. Munguia Tapia “Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors” M.I.T.; S.M. Thesis, 2003

[5] Kochenderfer, M.; Gupta, R. “Common Sense Data Acquisition for Indoor Mobile Robots” National Conference on A.I.; A.A.A.I., 2004

[6] Paradiso, J.; Lifton, J.; Feldmeier, M. “Plug” “http://www.media.mit.edu/resenv/classes/MAS961/plug/”; Media Lab, M.I.T. curriculum, 2005

[7] Ihler, A.; Fisher, J.; Moses, R.; Willsky, A. “Nonparametric Belief Propagation for Self-Calibration in Sensor Networks” Berkeley California; I.P.S.N., 2004

[8] Paskin, M.; Guestrin, C.; McFadden, J. “A robust architecture for distributed inference in sensor networks” Fourth International Conference on Information Processing in Sensor Networks ; IPSN, 2005

[9] Jepson, A.; Richards, W. “What makes a good feature?” Proceedings of the 1991 York Conference on Spatial Vision in Humans and Robots; Cambridge University Press, 1993

[10] Push Singh, Thomas Lin, Erik T. Mueller, Grace Lim, Travell Perkins and Wan Li Zhu “Open Mind Common Sense: Knowledge acquisition from the general public” Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems; Irvine, CA, 2002

[11] Hendrickson, B.; Leland, S. “The Chaco User’s Guide: Version 2.0” Sandia Tech Report; , 1995

[12] Fulford-Jones, T.; Wei, G.; Welsh, M. “A Portable, Low-Power, Wireless Two-Lead EKG System” Division of Applied Sciences, Harvard; Proceedings of the 26th Annual International Conference of the IEEE EMBS, 2004

[13] McFarland, D.; McCane, L.; Wolpaw, J. “EEG-Based Communication and Control: Short-term Role of Feedback” IEEE Transactions on Rehabilitation Engineering; Volume 6, No. 1, 1998

[14] Moore, M.; Kennedy, P. “Human Factors Issues in the Neural Signals Direct Brain-Computer Interface” Georgia State University; ASSETS, 2000