Fully Bayesian Human-Machine Data Fusion for Robust Dynamic Target Surveillance and Characterization *

Jeremy Muesing†, Luke Burks ‡, Michael Iuzzolino §, Danielle Albers Szafir¶, and Nisar Ahmed‖

I. Introduction

As US defense networks continue to expand automated ingestion and processing of high volume remote sensing data, human data analysts and system operators will always need to be kept in the loop, since sophisticated machine learning algorithms for automated event detection, tracking, and data fusion do not perform perfectly in all situations and decisions are ultimately made by people. A principal challenge is the end users of these automated systems (while generally highly trained and knowledgeable about many aspects of the problem domain and sensing assets) are not themselves algorithm experts or engineers and are unaware of algorithmic limitations. Current systems do not allow users to directly interact with these algorithms to mitigate issues, e.g., by updating prior target location or behavior information online for time-sensitive operations. New technologies that better balance automation (machine learning) and human oversight are needed to exploit the best of both worlds at various levels of the fusion pipeline. In contrast to current state of the art for deployable human-machine systems (where human interaction is very often treated in a post hoc manner), the next generation of automation must opportunistically leverage human reasoning abilities while still ensuring decision-making transparency and performance guarantees.

Our long-term research goal is to develop new fusion algorithms and interfaces that promote online collaborative human-machine perception for robust data analysis and fusion. The key idea is to allow analysts to communicate directly with automated machine learning algorithms via user friendly graphical interfaces for real-time information exchange and data visualization. Specifically, such algorithms and interfaces should (i) let analysts voluntarily push new information directly to automation, without exposing its inner workings and while accounting for uncertainties in human reporting; and (ii) let automation actively request useful information from analysts to boost performance via online querying. This allows the operator to act as an additional 'human sensor' capable of providing useful/timely information in situations that automated reasoning otherwise is unable to resolve satisfactorily on its own. These capabilities will require a combination of probabilistic machine learning, data fusion, and human input processing algorithms into a unified framework that enables bidirectional information exchange between automation and analysts for improved dynamic target analysis and decision-making under uncertainty. By using human reasoning to fill in automated machine perception/reasoning gaps, these efforts aim to measurably improve performance and robustness of state-of-the-art automated detection, tracking, and data fusion algorithms.

This paper looks at problem (i) (the information push) in the context of a generic large-scale dynamic multi-target characterization problem, where a single human operator monitors events and objects of interest over a large-scale surveillance area simultaneously scanned by multiple sensing assets in real time. In this general set up, the human operator interacts with an automated data fusion pipeline to supervise the event/object detection, tracking and labeling process (e.g., by confirming/rejecting particular tracks as objects of interest/false alarms or new target types at the 'front end' of the automated pipeline, adjusting sensing parameters in the 'back end', etc.), as well as to make follow-up decision recommendations for particular objects/events of interest in a timely manner. We focus on the problem of how target type information provided by the human operator in real-time can be fused with uncertain target type characterization information produced by the automated fusion pipeline. In particular, we consider how human operators can provide additional information in the form of positive or negative observations to enhance the target type probabilities produced by machine learning algorithms for each tracked object of interest in cases where the number of known possible target types is finite. We specifically address the problem of modeling the error characteristics of operator observations, while simultaneously fusing these with automated probabilistic assessments of target type for online target characterization.

*This work was funded by the US Air Force Space and Missile Systems Center, under contract BAA FA8810-17-C-0006.
†Graduate Research Assistant, Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado Boulder
‡Graduate Research Assistant, Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado Boulder
§Graduate Research Assistant, Computer Science, University of Colorado Boulder
¶Assistant Professor, Department of Information Science, University of Colorado Boulder
‖Assistant Professor, Ann and H.J. Smead Aerospace Engineering Sciences, University of Colorado Boulder, AIAA Member

To achieve this goal, we address two major technical challenges:

1) 'Stand-alone' human operator observation error characteristics are generally difficult and cumbersome to parameterize and calibrate *a priori*. For instance, when the space of possible target types is large, then any corresponding parameterization of human observation errors must grow to accommodate all the possible ways the human could mischaracterize multiple similar target types. Furthermore, unlike 'hard' data from conventional sensors (radar, EO/IR, visual, etc.), 'soft' data provided by humans can have highly unusual properties that, despite being information rich, make them difficult to process and combine with other information sources. For instance, error probabilities associated with human observations may not have stationary or even unique statistical properties (even once processed into a standardized format). As such, successive observations that are streamed from a single human are also not expected to be independent and identically distributed (i.i.d.) relative to one another, making error model parameter estimation even more challenging.
2) In real world scenarios, it is common for human reports to be dependent on other information that may already be present/fused elsewhere in a fusion pipeline. Without proper statistical modeling and blending of all information sources, this dependence leads to undesirable double-counting of information. For example, a human operator's target type observations may not be statistically independent of the target type assessments produced by the automated fusion pipeline.

We address these issues by developing a new fully Bayesian probabilistic model for Markovian online human-machine target characterization and data fusion which accounts for statistical dependencies between successive human observations as well as between probabilistic machine and conditionally dependent human target assessments. Our approach uses Dirichlet prior probability distributions over human model parameters to cope with parameter uncertainty and training/calibration data sparsity in practical settings, and uses approximate Gibbs sampling-based Bayesian inference to perform posthoc fusion of probabilistic machine assessments and human target type assessments. The use of a Dirichlet distribution allows for computationally fast and efficient online Gibbs sampling. This further enables easy scaling to scenarios requiring simultaneous error modeling and data fusion with multiple operators.

In the full version of this paper, we will provide an in-depth derivation and justification for our new probabilistic model and approximate Bayesian inference procedure for data fusion. We will also provide an extensive set of proof-of-concept results using synthetic target surveillance scenario data to demonstrate the main features of our approach, along with performance comparisons of our approach versus other fusion approaches, such as the current baseline fusion approach which does not use human input for tracking, as well as non-Bayesian maximum likelihood fusion. The remainder of this extended abstract highlights relevant concepts and previous work in the area of human-machine data fusion, and provides some additional details on the technical problem set-up, our novel Bayesian fusion model, and our approximate inference approach for online data fusion.

II. Background and Related Work

Probabilistic models and Bayesian reasoning provide a powerful general framework for augmenting automated reasoning and perception systems with 'soft data'—observations originating from human sources [1]—to complement 'hard data' from conventional sensors such as lidar, cameras, sonars, etc. in partially observable environments. For instance, human pilots/payload specialists in wilderness search and rescue missions can interpret video feeds and electro-optical/IR data streams provided by small unmanned aircraft, and can spot important clues that help narrow down probable lost victim locations and movements [2]. Likewise, in large-scale surveillance for defense applications, dismounted soldiers can provide evidence on the whereabouts and behaviors of potential intruders moving across unsecure areas. It is desirable to directly fuse such soft data with hard data from UAV patrols to improve intruder detection and tracking performance [3, 4]. Soft data integration also lets human supervisors of automated systems stay 'in the loop' without overloading them with cognitively demanding dynamic planning tasks [5].

A key problem is how should soft sensor data be formally integrated with hard data to augment automated algorithms? Soft data can be broadly related to either abstract phenomena that cannot be measured by robotic sensors (e.g. labels for object categories and behaviors) as well as measurable dynamical physical states that must be monitored constantly (object position, velocity, attitude, temperature, size, mass, etc.) [1]. This work focuses on the former, with the key assumption that *humans are **not** oracles* (where oracles have perfect knowledge). As with any other sensor data, human observations are subject to errors, limitations and ambiguities that must be modeled properly. We aim to adapt widely used statistical sensor fusion, state estimation and machine learning algorithms—Bayes filters, Hidden Markov models, and the like—so that soft data can be exploited with minimal effort on the part of the automated system and the human operator.

Fig. 1 Notional operator interface for simulated scenario with synthetic data.

Preliminary approaches to Bayesian modeling and fusion techniques allowed human sensors to directly interact [4, 6–10] with autonomous Bayesian state estimation and perception algorithms for dynamic target and event search, detection, localization, and tracking problems. The resulting fusion algorithms are grounded in probabilistic reasoning and models, and thus enable 'plug and play' functionality—i.e., human sensors can directly plug into filters and algorithms already used by automated systems for state estimation, and automated systems can continue to function as usual even in the absence of human input. Kaupp et al. [6] and Wang et al. [10] extend the core concept one step further to consider how uncertainties in human sensor model parameters can be accounted for in data fusion using maximum likelihood estimation, whereas Ahmed et al. [4] considers a fully Bayesian reasoning approach. Dani et al. [11], Mehta et al. [12], and Bishop et al. [13] have also considered alternative human-machine communication interfaces and probabilistic models for soft data fusion in dynamic target tracking problems. However, a key assumption for data modeling and fusion in all these works is that observations from a single observation remain i.i.d. and are further independent of other available information sources. Our approach does not assume i.i.d. to limit the double-counting of information and allows for the statistical dependencies between successive human observations.

III. Problem Setup and Proposed Solution Sketch

We consider how human operators can provide additional information in the form of positive/negative target type observations to enhance the target type probabilities produced by machine learning algorithms for each tracked object of interest in scenarios with a finite number of possible target types. In order to have a complete contextual picture for all tracks, the operator can access data associated with each track by visualizing various input/output data layers of the fusion pipeline, from raw sensor data feeds all the way up to automated track generation and labeling. However, for the purposes of the present work, each processing step of the fusion pipeline is effectively treated as a 'black box' whose inner workings cannot be accessed or altered by the operator. In this paper, we consider a case with only a finite number of tracks. Each target track is associated with a track ID, a mean state vector and covariance produced by a tracking Kalman filter. The associated label probabilities are displayed to the operator, and the raw/processed data for those tracks are accessible to the operator. Fig. 1 shows a notional example interface that we have built for a simulated tracking scenario with synthetic dynamic target data. The bottom right panel of the interface shows the input panel where operators can provide positive or negative observations to support/reject possible target classifications. Above this input panel is a bar graph representation for the automated pipeline's current probabilistic type assessment of a particular selected target track. In this synthetic example, there are 5 possible distinct target types for each possible track.

The data associated to each target has many different features that an operator can examine through the interface to make their assessment of target type, (e.g. location, velocity profiles, intensity data) although our approach could easily scale to more sensor data and features. We generated synthetic tracking data to approximate track detection and dynamic target profiles seen in a real world application scenario. For this paper, we generated five very similar profiles in order to model a scenario where the autonomy would require human assistance. Five signal return intensity profiles are shown in Figure 2. In addition to the basic intensity profile shapes, Gaussian noise is added as well as attenuation due to external/uncontrollable environmental effects on the sensors, e.g. due to weather.

Consider the problem of determining the true target type label X for a single selected target over a fixed frame of data gathered by sensors and provided by the human up until some time. Prior to a set of new observations being provided by the human, the automated fusion pipeline produces a prior probability $P(X)$ over the target type, e.g. using a Hidden Markov model [14] to associate raw and/or processed data signals obtained from the sensors to the most likely target profile according to some training database (other probabilistic algorithms could be used as well, provided they provide some measure of uncertainty on the final target type labels). Suppose also, the human then provides a series of N assessment observations $O_1, O_2, ..., O_{N-1}, O_N$ for a given data frame, where $O_i \in \{0^+, 1^+, ..., 5^+, 0^-, 1^-, ..., 5^-\}$, where 0^+ means 'selected target is of type 0', 0^- means 'selected target is not of type 0', 1^+ means 'selected target is of type 1', 1^- means 'selected target is not of type 1', etc. Then nominally we seek a Bayesian posterior distribution over the target type,

$$P(X|O_{1:N}) = \frac{P(X, O_{1:N})}{P(O_{1:N})} = \frac{P(O_{1:N}|X)P(X)}{P(O_{1:N})} \tag{1}$$

where the term $P(O_{1:N}|X)$ represents the joint data likelihood for the observed human data in the current data frame. The key challenge for fusion is to find a suitable representation for this data likelihood that accounts for dependencies between the observations (human target type assessments) in the sequence $O_{1:N} = O_1, ..., O_N$ as well as between $O_{1:N}$ and the automated pipeline's assessment X. The data likelihood thus represents the statistical error model for the human operator. In addition, we must contend with the fact that limited training data will be available to identify $P(O_{1:N}|X)$ with high confidence – and therefore the statistical error model for the human operator will itself also be uncertain.

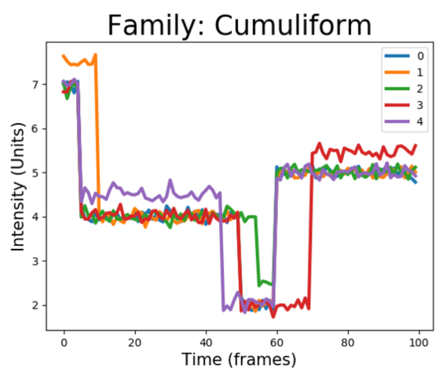

Fig. 2 Signal intensity profiles from simulated noisy sensor data for five different target types.

A. Graphical Model

To model the dependencies within $O_{1:N}$ and between the observations and X, we propose to use a Markov model to capture conditional dependencies between successive observations O_k and immediately preceding observations O_{k-1} as well as X. In this way, the joint observation likelihood term for a given data frame can be expressed as

$$P(O_{1:N}|X) = P(O_1|X) \prod_{j=1}^{N-1} P(O_{j+1}|O_j, X) = \theta_1 \prod_{j=1}^{N-1} \theta_{j+1}, \tag{2}$$

where we have defined θ_i as the model parameter which represents the ith conditional probability factor of the observation data likelihood under the Markov dependence model. To capture the uncertainty in the θ_i parameters in light of limited data, we can additionally impose a prior probability distribution on each θ_i. Consider, for instance, $N = 2$ observations where we have $\theta_1 = P(O_1|X)$ and $\theta_2 = P(O_2|O_1, X)$. Since the outcome space of the random variables X and O_1 and

O_2 are all discrete/categorical in nature, it follows that θ_1 and θ_2 represent conditional probability tables (CPTs) for multinomial distributions. Therefore, θ_1 and θ_2 can be modeled by arrays whose columns must sum to 1 for a particular configuration of conditioning variables. For instance, θ_1 must be a 10×5 array, where each column j represents the CPT for $P(O_1|X = j)$, and entry (m, j) of θ_1 represents the conditional probability $P(O_1 = m|X = j)$, for $m \in \{1, ..., 10\}$ (indexed 1-to-1 to the outcome space for O_1) and $j \in \{0, ..., 5\}$ (possible true target types). Likewise, θ_2 must be a $10 \times 10 \times 5$ array to represent the full CPT for $P(O_2|O_1, X)$. For $N > 2$, we impose a time invariant parameter tying assumption which forces $P(O_k|O_{k-1}, X)$ to be identical for any $k \geq 2$, so that $\theta_2 = \theta_3 = ...\theta_N$, thus greatly reducing the number of parameters required to model the human operator within a given frame.

If we assume suitable model parameter prior pdfs $p(\theta_1)$ and $p(\theta_2)$, and marginalize over θ_1 and θ_2 to account for our ignorance of the true values of these parameters, then the posterior Bayesian data fusion update for the target type probability becomes

$$P(X|O_{1:N}) = \frac{\int P(X, O_{1:N}, \theta_1, \theta_2) d\theta_1 d\theta_2}{P(O_{1:N})} \tag{3}$$

$$= \frac{\int P(O_{1:N}|X, \theta_1, \theta_2) P(X) p(\theta_1) p(\theta_2) d\theta_1 d\theta_2}{P(O_{1:N})} = \frac{\int \theta_1 \prod_{j=1}^{N-1} \theta_2 P(X) p(\theta_1) p(\theta_2) d\theta_1 d\theta_2}{P(O_{1:N})}, \tag{4}$$

where we have used the Markov observation model in the last expression and slightly abused notation to indicate that θ_1 and θ_2 represent the specific entries of the CPTs corresponding to the observations in $O_{1:N}$. In general, $P(X|O_{1:N})$ is analytically intractable to compute, to due to the complex conditional dependencies rendered by the observation sequence and the unknown prior model parameters on the target type X. This is also evident from the probabilistic graphical model of the data fusion problem as shown in Figure 3. Note that a variation of the data fusion problem is to simultaneously estimate both X and the unknown human model parameters, via the joint Bayesian posterior

$$P(X, \theta_1, \theta_2|O_{1:N}) = \frac{\theta_1 \prod_{j=1}^{N-1} \theta_2 P(X) p(\theta_1) p(\theta_2)}{P(O_{1:N})}, \tag{5}$$

which is also analytically intractable to compute. This latter posterior allows for display/assessment of operator reliability and performance through interpretation of the statistics of the parameters θ_1 and θ_2, which carry information about the operator's true positive, false positive, true negative and false negative rates for each possible target type.

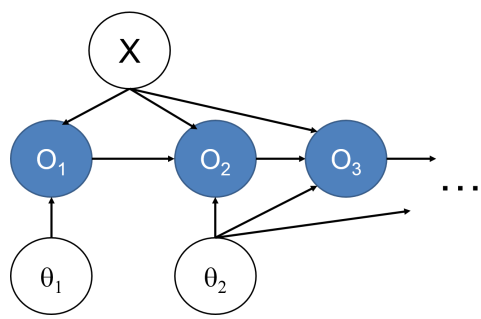

Fig. 3 Graphical model of the human-machine target type data fusion problem (observed variables shaded).

B. Dirichlet Priors and Gibbs Sampling

In the full paper, we will describe how to obtain an approximation to the above posterior distributions via a computationally efficient and fast Gibbs sampling procedure. We will describe how Dirichlet prior distributions can be used to define $p(\theta_1)$ and $p(\theta_2)$, and thus easily provide the conditional posteriors $P(X|\theta_1, \theta_2, O_{1:N})$ and $P(\theta_1, \theta_2|X, O_{1:N})$ required for Gibbs sampling. We will also describe how we set up these priors and describe performance sensitivity to these priors in different operating conditions. Our results with synthetic data will show that as updates are given by the human, these priors will lead to posteriors over the parameters that eventually conform to the actual values that describe individual operator, and lead to statistically consistent/robust values for the posterior over X, even as operators make mistakes and deal with multiple target tracks in succession (requiring an extended graphical model with additional versions of X and O variables for each target classification 'frame').

References

[1] Hall, D. L., and Jordan, J. M., *Human-centered Information Fusion*, Artech House, 2010.

[2] Goodrich, M. A., Morse, B. S., Engh, C., Cooper, J. L., and Adams, J. A., "Towards using Unmanned Aerial Vehicles (UAVs) in Wilderness Search and Rescue," *Interaction Studies*, Vol. 10, No. 3, 2009, pp. 453–478.

[3] Ahmed, N., Casbeer, D., Cao, Y., and Kingston, D., "Bayesian Hidden Markov models for UAV-enabled target localization on road networks with soft-hard data," *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VI*, Vol. 9464, International Society for Optics and Photonics, 2015, p. 94640Q.

[4] Ahmed, N., Campbell, M., Casbeer, D., Cao, Y., and Kingston, D., "Fully Bayesian Learning and Spatial Reasoning with Flexible Human Sensor Networks," *Proceedings of the 2015 Int'l Conf. on Cyberphysical Systems (ICCPS 2015)*, ACM/IEEE, 2015, pp. accepted, to appear.

[5] Lewis, M., Wang, H., Velgapudi, P., Scerri, P., and Sycara, K., "Using humans as sensors in robotic search," *12th International Conference on Information Fusion (FUSION 2009)*, Seattle, WA, 2009, pp. 1249–1256.

[6] Kaupp, T., Douillard, B., Ramos, F., Makarenko, A., and Upcroft, B., "Shared Environment Representation for a Human-Robot Team Performing Information Fusion," *Journal of Field Robotics*, Vol. 24, No. 11, 2007, pp. 911–942.

[7] Bourgault, F., Chokshi, A., Wang, J., Shah, D., Schoenberg, J., Iyer, R., Cedano, F., and Campbell, M., "Scalable Bayesian human-robot cooperation in mobile sensor networks," *International Conference on Intelligent Robots and Systems*, 2008, pp. 2342–2349.

[8] Kaupp, T., Makarenko, A., and Durrant-Whyte, H., "Human-robot communication for collaborative decision making: A probabilistic approach," *Robotics and Autonomous Systems*, Vol. 58, No. 5, 2010, pp. 444–456. doi:10.1016/j.robot.2010.02.003, URL `http://linkinghub.elsevier.com/retrieve/pii/S0921889010000400`.

[9] Ahmed, N., Sample, E., and Campbell, M., "Bayesian Multi-categorical Soft Data Fusion for Human-Robot Collaboration," *IEEE Transactions on Robotics*, Vol. 29, No. 1, 2013, pp. 189–206.

[10] Wang, D., Kaplan, L., Le, H., and Abdelzaher, T., "On truth discovery in social sensing: A maximum likelihood estimation approach," *Proc. of the 11th Int'l Conf. on Information Proc. in Sensor Networks*, ACM, 2012, pp. 233–244.

[11] Dani, A., McCourt, M., Curtis, J., and Mehta, S., "Information fusion in human-robot collaboration using neural network representation," *2014 IEEE Int'l Conf. on Systems, Man and Cybernetics*, IEEE, 2014, pp. 2114–2120.

[12] Mehta, S., McCourt, M., Doucette, E., and Curtis, J., "A touch interface for soft data modeling in Bayesian estimation," *2014 IEEE Int'l Conf. on Systems, Man and Cybernetics*, IEEE, 2014, pp. 3732–3737.

[13] Bishop, A. N., and Ristic, B., "Fusion of natural language propositions: Bayesian random set framework," *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, IEEE, 2011, pp. 1–8.

[14] Bishop, C., *Pattern Recognition and Machine Learning*, Springer, New York, 2006.