Identity Uncertainty

Stuart Russell
Computer Science Division
University of California, Berkeley, CA 94720, USA
russell@cs.berkeley.edu

Abstract

We are often uncertain about the identity of objects. This phenomenon appears in theories of object persistence in early childhood; in the well-known Morning Star/Evening Star example; in tracking and data association systems for radar; in security systems based on personal identification; and in many aspects of our everyday lives. The paper presents a formal probabilistic approach to reasoning about identity under uncertainty in the framework of first-order logics of probability, with application to wide-area freeway traffic monitoring.

1 Introduction

Object identification—the task of deciding that two observed objects are in fact one and the same object—is a fundamental requirement for any situated agent that reasons about individuals. The aim of this paper is to explore uncertain reasoning about object identity, with the ultimate aim of understanding how internal representations of individual objects should be created and manipulated within a complete intelligent agent.

The existence of individuals is central to our conceptualization of the world. While *object recognition* deals with assigning objects to categories, such as 1988 Toyota Celicas or adult humans, *object identification* deals with recognizing specific individuals, such as one's car or one's spouse. One can have specific relations to individuals, such as ownership or marriage. Hence, it is often important to be fairly certain about the identity of the particular objects one encounters.

Formally speaking, identity is expressed by the equality operator of first-order logic. Having detected an object C in a parking lot, one might be interested in whether $C = MyCar$. Because mistaken identity is always a possibility, this becomes a question of the *probability of identity*: $P(C = MyCar|$ all available evidence).

For many years, philosophers have pondered this question (or rather, more general versions thereof). The famous example of the Evening Star and Morning Star, long thought to be distinct but in fact the same object (Venus), shows that one can be mistaken about identity and that it is really a matter of inference from accumulated evidence. The doctrine of the Indiscernibility of Identicals states, uncontroversially, that if $a = b$, $F(a) \Leftrightarrow F(b)$ for any predication F. The Identity of Indiscernibles states the converse, thereby pointing out that identity, and hence the choice of what one considers to be individuals, is strongly related to the choice of predications one wishes to make.

There has been little analytical work on object identity in AI or, it seems, in probability and statistics. The one area where the issue has been addressed is that of tracking and data association, as we discuss in Section 5. This paper sketches a more general framework, in which the data association model can be accommodated and extended. It draws extensively on the previous work of the author and his colleagues and students [6, 10, 12, 13, 9].

We begin in Section 2 with what is perhaps the simplest possible scenario involving identity uncertainty, and we apply straightforward mathematical techniques to solve it. Section 3 developes a formal language, combining probability and first-order logic, that allows identity questions to be posed in general settings with many objects, properties, and relations. Section 4 shows how inference for this formal language may be achieved using Markov chain Monte Carlo (MCMC) algorithms, where the Markov chain is defined on a state space consisting of the relational models of the first-order language. Section 5 shows how MCMC may be used for reasoning about vehicle identities in the context of a large-scale traffic surveillance system.

2 Balls and Urns

Papers on the foundations of probability are fond of drawing balls from urns. Here, we consider the following scenario. The experimenter is given an urn containing an unknown collection of balls, and k are performed. In each, a ball is removed from the urn, its colour is examined, and it is returned to the urn, which is then shaken vigorously. The standard question in the literature is, "What is the probability that the next ball drawn is red?" Instead, we will ask, "How many balls are in the urn?"

Intuitively, it is clear that reasoning about identity is required in order to answer this question. Consider the case where the experimenter can determine that the first five balls drawn are identical—that is, the *same* ball is drawn five times. Then it seems reasonable to infer that the urn probably contains just one ball.

A formal probability model for this problem is defined on the following random variables:

- N is the number of balls, with prior $P(N)$.
- $\mathbf{C} = C_1, \ldots, C_N$ are the true colours of the N balls in the urn, each with an identical prior $P(C_i)$.
- $\omega = \omega_1, \ldots, \omega_k$ are the true identities of the balls drawn; the value of each is an integer in $\{1, \ldots, N\}$. By assumption, these variables are independent and uniformly distributed.
- $\mathbf{Y} = Y_1, \ldots, Y_k$ are the observed colours of the k balls drawn, each distributed according to an identical sensor model $P(Y_j | C_{\omega_j})$.

The joint distribution is then

$$P(N, \mathbf{C}, \omega, \mathbf{Y}) = P(N) \left(\prod_{i=1}^{N} P(C_i)\right) \left(\tfrac{1}{N}\right)^k \prod_{j=1}^{k} P(Y_j | C_{\omega_j}) \quad (1)$$

Notice how the identity of the balls enters into this expression, by selecting the true colours of the observed balls and hence correlating the observed colours.

The question, "How many balls are there?" is answered by the posterior distribution $P(N | \mathbf{Y} = \mathbf{y})$, given by

$$\alpha P(N) \left(\frac{1}{N}\right)^k \sum_{\mathbf{c}} \left(\prod_{i=1}^{N} P(C_i)\right) \sum_{\omega} \prod_{j=1}^{k} P(Y_j | C_{\omega_j}) \quad (2)$$

(In this expression, the sum over $\mathbf{c}$ may be replaced by an integral if the colour distribution is continuous.) When $k < N$, this expression can be simplified somewhat, but in general the summation over all N^k identity assignments must be calculated. This is characteristic of all the identity uncertainty problems we have studied.

One question of particular interest is whether the number of balls can be identified exactly in the limit of many experiments, i.e., as $k \to \infty$. The answer to this question depends on the nature of the colour distribution and sensor model. If there can be indistinguishable but non-identical balls—for example, if the colour distribution is discrete and the sensor model is exact—then it is impossible for any set of observations to distinguish among the situations with N, $2N$, $3N$, ... balls. In that case, the posterior will tend to a collection of delta functions placed at values of N equal to multiples of the lowest common multiple of the denominators of the reduced-form fractions of each colour. For example, if the observed fractions of three colours in the limit are 1/3, 1/3, and 1/3, then there could be 3, 6, 9, ... balls. On the other hand, if the colour distribution is continuous, indistinguishable balls occur with probability zero and the posterior converges to the true number as $k \to \infty$.

3 A formal theory

The preceding section gave an example and developed a specific formula for that example. In this section, we describe a general representation language that combines probability theory with a restricted fragment of first-order logic, sufficient to represent many cases of identity uncertainty. Given this language, a single algorithm can be used to answer questions about all such cases.

The basic principles of first-order probabilistic logic (FOPL) were given by [5]. Each model structure of a FOPL knowledge base should be viewed as a probability measure over the possible worlds (logical models) defined by the constant, function, and predicate symbols of the knowledge base. Entailment between probabilistic assertions is then defined identically to ordinary logical entailment, i.e., via truth in all model stuctures. Given a complete knowledge base—one that fixes a unique model structure—the probability of a sentence is defined to be the sum of probabilities assigned to all the possible worlds in which that sentence is true.

We extend the relational probability models of [8] as follows:

- A set $\mathcal{C}$ of *classes* denoting sets of objects, related by subclass/superclass relations. Each class has an associated prior distribution over the cardinality of the class.
- A set $\mathcal{I}$ of *named instances* denoting objects, each an instance of one class.

- A set $\mathcal{B}$ of *simple attributes* denoting functions. Each simple attribute B has a domain type $Dom[B] \in \mathcal{C}$ and a range that is a finite, enumerated set of values $Val[B]$.

- A set $\mathcal{A}$ of *complex attributes* denoting relations. Each complex attribute A has a domain type $Dom[A] \in \mathcal{C}$ and a range type $Range[A] \in \mathcal{C}$. We allow for *reference uncertainty*, i.e., the values of each complex attribute may be unknown.

- With each complex attribute A, we associate a simple *reference attribute* $ref[A]$, such that $Val[ref[A]]$ is a finite, enumerated set of named instances. (Where no confusion arises, we may drop the $ref[\cdot]$ and use the attribute name itself.)

- A special relational attribute "=", standing for identity in the standard logical sense.

- A set of conditional probability models $P(B|Pa[B])$ for the simple attributes. A *parent chain* for a simple attribute is a nonempty chain of (appropriately typed) attributes $\sigma = A_1.\cdots.A_n.B'$, where B' is a simple attribute. $Pa[B]$ is the set of B's *parents*, namely the set of all instance variables reached by any parent chain for all possible combinations of values of all the uncertain complex attributes, as well as all the reference variables for those attributes.

- Dependencies are expressed as before by a conditional distribution $P(ref[A]|Pa[ref[A]])$. The parents of $ref[A]$ are those attributes or attribute chains that influence the choice of an instance as the value of attribute A.

Most applications of FOPL assume that there is no reference uncertainty, and that every instance is distinct (unique names assumption, hence no identity uncertainty). Thus, a possible world is defined by the values of the *instance variables*—the simple attributes for all named instances—and the knowledge base specifies a complete distribution over all possible worlds. With reference uncertainty, possible worlds also vary according to the relations that hold among objects, which determine in turn the probabilistic influences among instance variables. With identity uncertainty, we also must specify which instances in the representation map to which objects in the possible world, and how many objects the world contains—thus, the possible world is a complete structure and interpretation as in first-order predicate calculus with equality. This means that the probability of any given sentence is given by summing over all possible identity relations and numbers of objects, all possible relational structures among objects, and all possible values for the instance variables.

To represent the balls-and-urn example, we need two classes, $Ball$ and $Observation$. A prior is specified for the cardinality of the $Ball$ class. There is one complex attribute, $generated\_by$, which maps an $Observation$ to the $Vehicle$ that generated it. Each $Observation$ of a vehicle reports a $colour$, which depends (through the sensor model) on the $colour$ of the corresponding $Ball$. For each observation, a new $Observation$ instance is added to the KB with its colour set to the measured value, and with a new $Ball$ instance attached. A query can be issued to the inference algorithm asking for the posterior cardinality of the $Ball$ class.

4 MCMC on logical models

Exact inference for full FOPL languages is undecidable; for decidable fragments with reference uncertainty, the complexity is very high. Initial experiments suggest, however, that approximate inference using MCMC is a promising approach [9].

MCMC [4] generates samples from a posterior distribution $\pi(x)$ over possible worlds x by defining a Markov chain whose states are the worlds x and whose *stationary distribution* is $\pi(x)$. In the *Metropolis–Hastings* method (henceforth M-H), transitions in the Markov chain are constructed in two steps:

- Given the current state x, a candidate next state is generated from the *proposal distribution* $q(x'|x)$, which may be (more or less) arbitrary.

- The transition to x' is not automatic, but occurs with an *acceptance probability* defined by

$$\alpha(x'|x) = min\left(1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\right)$$

It is not necessary that all the variables of state x be updated simultaneously, in a single transition function. *Single-component* M-H alters each variable in turn. It is also possible to factor q into separate transition functions for various subsets of variables. Provided that q is defined in such a way that the chain is ergodic, this transition mechanism defines a Markov chain whose stationary distribution is $\pi(x)$.

The Gibbs sampling algorithm for Bayesian networks [11] is a special case of Metropolis–Hastings in which the proposal distribution samples a single variable X_i using the distribution $P(X_i|\mathbf{mb}(X_i))$, where $\mathbf{mb}(X_i)$ denotes the current values of the variables in the Markov blanket of X_i (its parents $Pa[X_i]$, children Y_j, and children's other parents). In this case, the acceptance probability is always 1. One can show easily that

$$P(X_i|\mathbf{mb}(X_i)) = \alpha P(X_i|Pa[X_i])) \prod_j P(Y_j|Pa[Y_j]) \tag{3}$$

Gibbs sampling is very simple and also *local*: transitions are generated referring only to parts of the model directly connected to the variable in question. Hence, *the cost per transition is typically independent of model size*. M-H sampling is also typically local because all the parts of the model that are not changed by the transition cancel in the ratio $\pi(x')/\pi(x)$. In particular, if the proposal concerns a single variable X_i, this ratio reduces to $P(x_i'|\mathbf{mb}(X_i))/P(x_i|\mathbf{mb}(X_i))$, where x_i' is the proposed value of X_i and x_i is its current value. The M-H algorithm, unlike Gibbs, has the added advantage that the transition may often be computed without referring to the other values of X_i at all, as we will see.

MCMC can be applied to FOPL inference by definng the state space to be the set of possible worlds, as described above. For the balls-and-urn example, transitions may modify the true colours of the balls, the identities of the sampled balls, or the number of balls. For cases where the true posterior converges to a single delta function as $k \to \infty$, MCMC mixes well and the aprpoximation is very accurate. (It should be possible to give good convergence bounds for this case.) For cases with ambiguous posteriors, MCMC mixing with a simple "random walk" proposal on N will not converge very quickly because, for large k, the chain will become trapped in one of the peaks in teh posterior distribution. At present it is not known if the approximation problem is intractable or whether a better proposal can be designed.

5 Traffic surveillance

We have developed one large-scale application of MCMC for identity uncertainty—a traffic surveillance system that uses multiple cameras to asses the state of a large freeway network [10]. The sensors used in this project are video cameras placed on poles beside the freeway (Figure 1). Observations made at multiple cameras must be combined to track vehicles through the freeway network. Object identification is required for two purposes: first, to measure *link travel time*—the actual time taken for vehicles to travel between two fixed points on the freeway network; and second, to provide *origin/destination* (O/D) counts—the total number of vehicles traveling between any two points on the network in a given time interval.

Figure 1. Images from two surveillance cameras roughly two miles apart on Highway 99 in Sacramento, California. The top image is from the upstream camera, and the bottom image is from the downstream camera. Are the two boxed vehicles the same?

Combining multiple observations into tracks is a task addressed by the *data association* field—see [2, 1]. The intractability of exact data association inference [3] has led to many approximate methods. In our approach, MCMC is used to sample from possible track histories for all vehicles. MCMC transitions simply switch two tracks at some point in their histories. For two cameras, this gives a polynomial-time approximation [7]—the first approximation result for data association. Our experiments show that MCMC converges quickly—in 100 samples or so—as each new vehicle is added, even in state spaces with thousands of vehicles.

Figure 2 shows a simulated freeway network. The aim is to estimate the origin-destination counts between the two entry points and the three exit points. This requires tracking each object across the entire network. In this task, we compare two algorithms: Huang–Russell [6] and MCMC. The Huang–Russell algorithm gives a good approximation to pairwise identity probabilities, and then concatenates these identities to track vehicles over several cameras. Because MCMC does joint es-

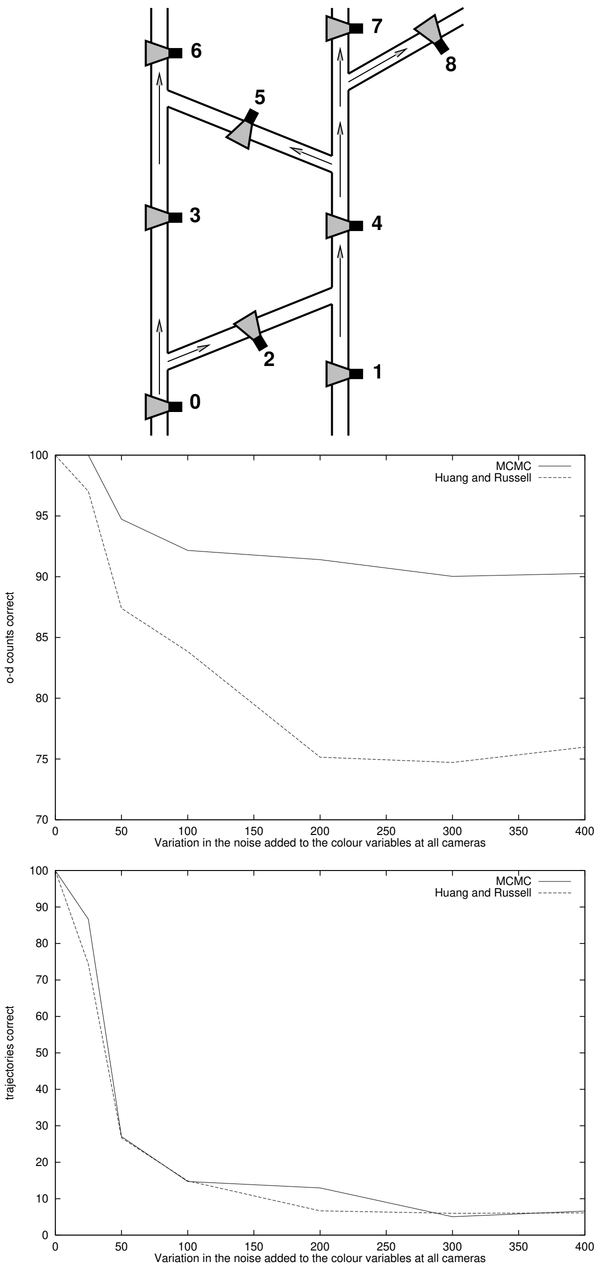

Figure 2. Top: Schematic diagram of simulated freeway network with nine cameras. Middle and Bottom: The average accuracy, measured over entrances and exits, of the origin-destination counts and vehicle matches for the MCMC and Huang–Russell algorithms, as a function of the colour noise variance.

timation over the whole network, we expect it to perform much better than Huang–Russell as the measurement noise increases. The results in Figure 2 bear this out. The second graph in the figure also shows that both methods are unable to find exact matches accurately for high levels of noise. The ability of the MCMC algorithm to recover reasonable counts despite the failure of individual matches suggests that its samples contain a reasonable amount of information about the ensemble behaviour of the vehicles.

6 Conclusions

This paper has introduced the general problem of identity uncertainty as a subject of study for AI and soft computing generally. MCMC was proposed as a promising technique that can be applied to perform approximate inference on general first-order probabilistic logics, by sampling from possible worlds. Many open problems remain. Perhaps the most important is that of finding natural ways to specify distributions over possible identity relations—i.e., how many things are there? How frequently do new things appear? Resolving this issue and combining the solution with existing methods—such as Bayes nets and relational probability models—should open up many new applications.

References

[1] Y. Bar-Shalom, editor. *Multitarget-multisensor tracking: Advanced applications*. Artech House, Norwood, Massachusetts, 1992.

[2] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, New York, 1988.

[3] I. J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10:53–66, 1993.

[4] W. Gilks, S. Richardson, and D. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman and Hall, London, 1996.

[5] J. Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46(3):311–350, 1990.

[6] T. Huang and S. Russell. Object identification in a Bayesian context. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, Nagoya, Japan, Aug. 1997. Morgan Kaufmann.

[7] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method. In D. S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*. PWS Publishing, Boston, 1997.

[8] D. Koller and A. Pfeffer. Semantics and inference for recursive probability models. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-00)*, Austin, Texas, July 2000. AAAI Press.

[9] H. Pasula and S. Russell. Approximate inference for first-order probab ilistic languages. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, Seattle, Aug. 2001. Morgan Kaufmann.

[10] H. Pasula, S. Russell, M. Ostland, and Y. Ritov. Tracking many objects with many sensors. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*, Stockholm, Sweden, Aug. 1999. Morgan Kaufmann.

[11] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[12] S. Russell. Expressive probability models for speech recognition and understanding. In *Proc. of the International Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Keystone, Colorado, Dec. 1999. IEEE.

[13] S. Russell. Expressive probability models in science. In *Proc. of the 2nd Int'l Conf. on Discovery Science*, Tokyo, Japan, Dec. 1999. Springer Verlag.