Int J Multimed Info Retr (2012) 1:249–262
DOI 10.1007/s13735-012-0021-5

REGULAR PAPER

Acquisition of multimedia ontology: an application in preservation of cultural heritage

Anupama Mallik · Santanu Chaudhury

Received: 1 August 2012 / Accepted: 9 September 2012 / Published online: 17 October 2012

Abstract A domain-specific ontology models a specific domain or part of the world. In fact, ontologies have proven to be an excellent medium for capturingpagebreak the knowledge of a domain. We propose an ontology learning scheme in this paper which combines standard multimedia analysis techniques with knowledge drawn from conceptual metadata to learn a domain-specific multimedia ontology from a set of annotated examples. A standard machine-learning algorithm that learns structure and parameters of a Bayesian network is extended to include media observables in the learning. An expert group provides domain knowledge to construct a basic ontology of the domain as well as to annotate a set of training videos. These annotations help derive the associations between high-level semantic concepts of the domain and low-level media features. We construct a more robust and refined version of the basic ontology by learning from this set of conceptually annotated data. We show an application of our ontology-based framework for exploration of multimedia content, in the field of cultural heritage preservation. By constructing an ontology for the cultural heritage domain of Indian classical dance, and by offering an application for semantic annotation of the heritage collection of Indian dance videos, we demonstrate the efficacy of ou approach.

Keywords Ontology learning · Multimedia ontology · MOWL · Video annotation

A. Mallik (✉) · S. Chaudhury
IIT, New Delhi, India
e-mail: ansimal@gmail.com

S. Chaudhury
e-mail: schaudhury@ee.iitd.ac.in

1 Introduction

An ontology is a "formal, explicit specification of a shared conceptualisation".[1] In other words, it is the formal representation of a set of concepts within a domain and the relationships between those concepts. It provides a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist, and their properties and relations. It is used to reason about the properties of that domain, and may be used to define the domain. Thus a domain ontology (or domain-specific ontology) models a specific domain, or part of the world. In fact, ontologies have proved to be an excellent medium for capturing the knowledge of a domain. In this paper, we propose a novel ontology learning scheme which utilizes domain experts' knowledge, combined with annotated examples of the domain to construct a multimedia ontology for effective use in retrieval applications. Ontologies have been used in multimedia retrieval applications [10,14], but applying ontology learning to improve multimedia retrieval, specially attuned to probabilistic reasoning as is required with multimedia data and linked observations, has not been attempted before.

Ontology construction is necessarily an iterative process. An ontology representing concepts and relationships of the domain can be constructed manually with a domain expert providing the inputs. In this process, there is a possibility of missing out some concepts and relations which may exist in the real-world, while coding some extra knowledge which might be obsolete. It is highly effective to fine-tune the knowledge obtained from the expert by applying learning from real-world examples belonging to the domain. An ontology refined in this manner is a better structured, logically valid

[1] Wikipedia definition.

model of the domain that it represents. The goal of ontology learning, thus, is to (semi-)automatically generate relevant concepts and relations from a given corpus and expert inputs.

The objective of our work is to devise a framework for learning a domain-specific ontology from multimedia data belonging to the domain, in order to provide a highly effective content-based access to the data contained in the repository. The novelty of our work is the ability to encode the highly specialized knowledge that experts of a scholarly domain have, into an ontological representation of the domain, and refine this knowledge by learning from observables in the multimedia examples of the domain. Combination of domain knowledge with example-driven supervised learning for generation of domain ontology for multimedia retrieval is a unique contribution of this work. Learning the ontology in our scheme employs the use of the Multimedia Web Ontology Language (MOWL) and its unique probabilistic reasoning framework for representing the multimedia ontology. MOWL representation allows a Bayesian network representation of the ontology snippets, and thus allows us to extend a standard Bayesian network learning algorithm for learning the structure and parameters of the multimedia ontology. We have shown the success of our technique by applying our work to a cultural heritage domain of Indian classical dance. We use the ontology-specified knowledge for recognizing concepts relevant to a video to annotate fresh additions to the video database with relevant concepts in the ontology.

2 Related work

Research work in ontology-based multimedia information retrieval (MIR) elaborates on *how* to use ontology for MIR but not on how to relate the ontology to multimedia data. For example, learning has been used in the LSCOM [15], but only for concept-detection, not for learning of ontology. Ontology learning refers to the automatic discovery and creation of ontological knowledge using machine-learning techniques, with little human intervention. A lot of research in ontology learning is happening but not in the multimedia domain. State-of-art ontology learning approaches have been discussed in [24]. According to this review, text is the most used medium for learning ontologies. Limited research exists in the area of ontology learning with multimedia content. In [25], the authors discuss the challenge of developing domain ontologies, specially for under-developed domains, which have no structural resources in existence. They propose the ROD methodology that can automatically discover concepts and relations from large-scale semi-structured and/or unstructured textual resources. An example of rule-based ontology learning can be found in the OntoLearn system [16], which extracts relevant domain terms from a corpus of text, relates them to appropriate concepts in a general-purpose ontology, and detects taxonomic and other semantic relations among the concepts. Amongst multimedia applications that use ontology learning, [11] presents a concept hierarchy of actions and propose a method for describing human activities from video images based on this hierarchy to generate a natural language sequence from a video sequence.

2.1 Bayesian learning

Bayesian learning is a common statistical machine-learning approach. Its use in ontology learning is limited by the lack of support in standard ontology languages like OWL for probabilistic reasoning. In [5], Ding et al. have proposed a probabilistic extension to OWL by using Bayesian networks, but this is limited to textual data. Here, we mention some of the research happening in the field of Bayesian network learning.

Starting from his tutorial on learning Bayesian networks in 1995 [9], Heckerman has published several works in this field. His research focuses on structural as well as parameter learning in Bayesian networks. Other algorithms and methods of structure learning in probabilistic networks include so-called *naive* Bayesian network learning, which states that classification is an optimal method of supervised learning in a Bayesian network if the values of the attributes of an example are independent given the class of the example. In [23], Zheng et al. have considered an extension of naive Bayes, where a subset of the attribute values is considered, assuming independence among the remaining attributes. Niculescu et al. [18] have used parameter constraints to learn the Bayesian network. Ramachandran et al. [19] use the back-propagation approach of the neural network to the Bayesian network learning. In [2], the authors have focused on the problem of learning probabilistic networks with known structure and hidden variables from data, defining what they call the Adaptive Probabilistic Network (APN) algorithm. They mention that an improvement for parametric learning algorithms like APN could be to allow a domain expert to pre-specify constraints on the conditional distributions. Buntine [4] gives a literature review discussing different methods of learning Bayesian networks from data.

Bayesian learning has been used in several applications of information retrieval. Neuman et al. [17] have described a model of IR based on Bayesian networks in [17]. In [1], we see the usage of Bayesian learning for Neural Networks in predicting both the location and next service for a mobile user movement. In [21], Town et al. have described how an ontology consisting of a ground truth schema and a set of annotated training sequences can be used to train the structure and parameters of Bayesian networks for event recognition. They have applied these techniques to a visual surveillance problem, and use visual content descriptors to infer high-level event and scenario properties. These applications work

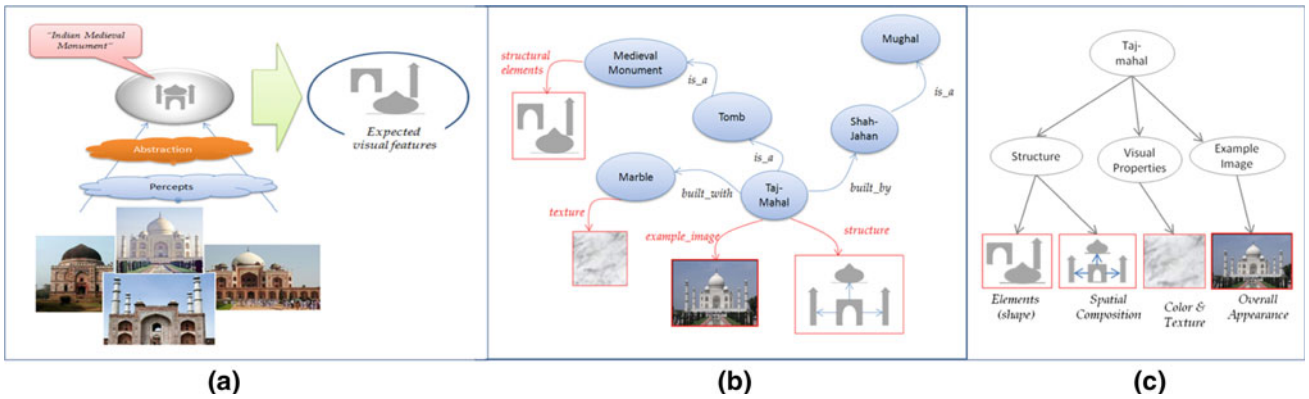

Fig. 1 **a** Perceptual modelling, **b** multimedia ontology of Indian monuments, **c** observation model of Taj Mahal

in generic, open domains where domain knowledge is not specialized, and there is no learning from meta-data attached to the videos.

Motivated by the developments in Bayesian learning, we have proposed in this paper, a scheme for learning a multimedia ontology encoded as a probabilistic Bayesian network. We detail our scheme for building and learning of a multimedia ontology for an example domain of Indian classical dance (ICD), verifying with experiments how the snippets of ontology learnt through our framework, help in more effective retrieval. The rest of this paper is organized as follows. Section 3 details the multimedia ontology representation scheme offered by the Multimedia Web Ontology Language. In Sect. 4, we give an overview of the ontology-based framework for a multimedia content management system, which uses our ontology learning scheme to build the ontology required for its working. Section 5 gives the details of our ontology learning scheme explaining how the ontology is learnt from domain experts' knowledge and labelled multimedia data. Section 6 gives details of an application of our ontology learning framework in learning a multimedia ontology for the heritage domain of Indian classical dance. Section 7 concludes the paper by summarizing our findings.

3 Multimedia ontology representation through MOWL

We have used the Multimedia Web Ontology Language [8] for representing the multimedia ontologies used in our experiments. An ontology encoded in a traditional ontology language, e.g. OWL, uses text to express the domain concepts and the properties. Thus, it is quite straightforward to apply such an ontology for semantic text processing. Semantic processing of multimedia data, however, calls for ontology primitives that enable modelling of domain concepts with their observable media properties. This kind of modelling is called **Perceptual Modelling**, an example of which is shown in Fig. 1a. Such modelling needs to encode the inherent uncertainties associated with media properties of concepts too. Traditional ontology languages do not support these capabilities.

In order to support semantic media processing, we use the ontology representation scheme offered by MOWL, that enables encoding of media properties for the concepts in a closed domain. The basic premise of MOWL is a causal model of the world, where real-world concepts (and events) lead to manifestation of media features in multimedia documents. This causal modelling distinguishes MOWL from OWL and other knowledge representation languages. The causal model can be used for abductive reasoning for concept-recognition in multimedia data, where the observed media features in a multimedia document can be *causally* explained as manifestations of concepts. Syntactically, MOWL is an extension of OWL. However, it supports probabilistic reasoning with *observation models* of the concepts, which can be interpreted as Bayesian networks with CPTs. This is in contrast to crisp Description Logic-based reasoning with traditional ontology languages. MOWL allows encoding of uncertainties which exist in the observation of multimedia data, and in some relations between concepts which are probabilistic. These can be specified as joint probabilities of a concept in relation with several other concepts. This kind of reasoning is useful in concept discovery in documents belonging to multimedia collections.

We have used MOWL to encode our domain ontology. MOWL provides the following functionality for a multimedia ontology representation:

- **Concepts and media properties**
  MOWL distinguishes between two types of entities, namely (a) the *concepts* that represent the real-world objects or events and (b) the *media objects* that represent manifestation of concepts in different media forms. Detection of the media objects leads to concept-recognition. For example, as shown in Fig. 1b, while a monument is a real-world concept, the visual image of the monument is a media object which represents its

media manifestation. As another example, a specific performance of a dance piece can be recognized by a set of gestures, postures and actions, which form a set of media objects representing the possible media manifestations for the dance performance of a particular conceptual category.

- **MOWL relations**
  Relations between the concepts play an important role in concept-recognition. For example, an important cue to the recognition of a medieval monument can be the visual properties of the stone it is built with (as shown in Fig. 1b). As another example, a classical dance form is generally accompanied by a specific form of music. Thus, detection of media properties characterizing the music form is an important cue to recognition of the dance form. In order to enable such reasoning, MOWL allows definition of a class of relations that imply "propagation" of media properties.

- **Specifying spatio-temporal relations**
  Complex media *events* can be defined in MOWL with constituent media objects and their spatio-temporal relations with formal semantics which is consistent with and can be executed with an extended MPEG-7 Query Engine proposed in [22]. For e.g., in a classical dance, a certain dance step is a choreographical sequence of certain dance postures. A multimedia ontology should be able to specify such concepts in terms of spatial/temporal relations between the components. MOWL defines a subclass of media objects called `<mowl:ComplexObject>` which represents composition of media objects related through spatial or temporal relations. Every complex object is defined by a spatial or temporal relation or *predicate* and two media objects—one the *subject* of the predicate relation and the other the *object* of the predicate. For example, a soccer goal can be represented as a complex object with *subject* "ball", *spatial predicate* "inside" and *object* "goalpost".

- **Uncertainty specification**
  The relations that associate concepts and media objects are causal relations and are generally uncertain in nature. For example, though certain gestures and postures are integral parts of a classical dance performance, they may be omitted or modified in a particular instance of a performance. Thus, these associations are probabilistic in nature. MOWL provides for specification of uncertainty of these associations in a multimedia domain by providing special constructs for defining Conditional Probability Tables (CPTs) and associating them with concepts and media objects.

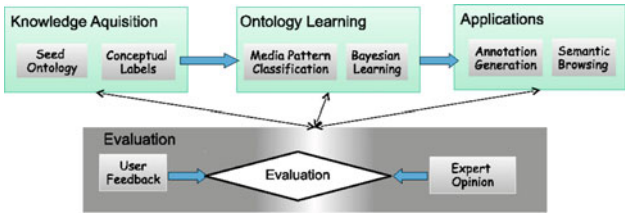

Fig. 2 Framework for ontology-based management of multimedia content

- **Reasoning with Bayesian networks**
  The knowledge available in a MOWL ontology is used to construct an observation model (OM) for a concept, which is in turn used for concept-recognition. This requires two stages of reasoning:

  1. *Reasoning for derivation of observation model for a concept.* This requires exploring the neighbourhood of a concept and collating media properties of neighbouring concepts, wherever media property propagation is implied. The resultant observation model of a concept is organized as a Bayesian network (BN).

  2. *Reasoning for concept-recognition.* Once an observation model for a concept is created, it can be used for concept recognition. We use an abductive reasoning scheme that exploits the causal relations captured in the observation model (Fig. 1c).

4 Ontology-based management of multimedia resources

Multimedia resources pertaining to a domain can include digital replicas of domain artefacts, events, etc. Depending on the domain, these can be videos or still images of soccer matches, paintings, sculpture or dance performances, as well as the contextual knowledge relating to these resources, which is contributed by domain experts. Our proposed scheme for multimedia resource management is motivated by the need for relating the digital objects with contextual knowledge, to make the former more usable. With these requirements, we have proposed an ontology-based framework for multimedia content management with flexible structure and dynamic updation. In this paper, we focus on our technique of *building* the multimedia ontology, which is the backbone of this framework, with the help of knowledge obtained from domain experts and then learning it from real-world data which is part of the digital resources of the domain. There are four main stages in the framework, namely **knowledge acquisition**, **ontology learning**, **application** and **evaluation**, as shown in Fig. 2.

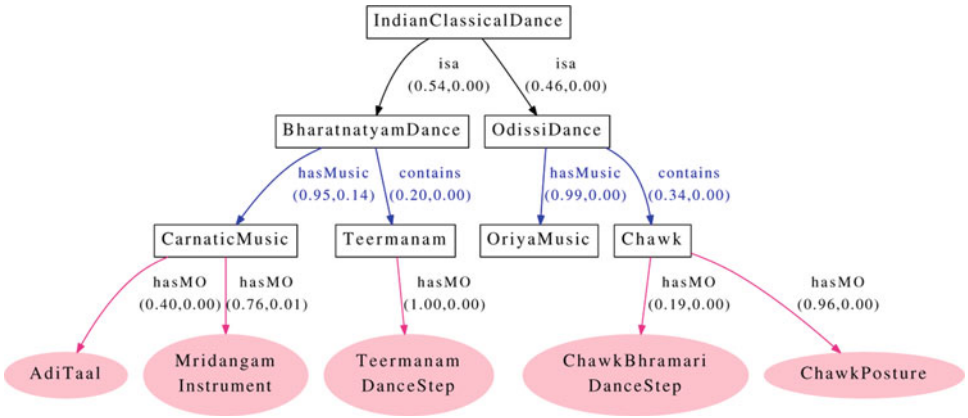

Fig. 3 Basic Ontology snippet Γ_B of the ICD domain as specified by the domain experts, enriched with multimedia

1. **Knowledge acquisition** This stage deals with acquiring the highly specialized knowledge of a domain and encoding it in a domain-specific ontology. It also involves collecting the multimedia data of the domain and building a digital collection. To begin with, a basic *s*eed ontology for the domain is hand-crafted by a group of domain experts. The ontology includes the domain concepts, their properties and their relations. The domain experts also provide *c*onceptual labels to a training set of multimedia data. They annotate the multimedia files and their segments, based on their observations, in such a way that the labels correspond to domain concepts in the ontology.
2. **Ontology learning** At this stage, the basic ontology, enriched with multimedia data, is further refined and fine-tuned by applying machine-learning from the training set of labelled data. We use MOWL to represent the ontological concepts and the uncertainties inherent in their media-specific relations. The multimedia ontology thus created, encodes the experts' perspective and needs adjustments to attune it to the real-world data. Conceptual annotations help build the case data used for applying a machine-learning technique called the Full Bayesian network (FBN) learning to refine the ontology. An important part of the ontology learning stage is the development of **media pattern classifiers** which can detect media patterns corresponding to lowest-level media nodes in the ontology based on the presence of content-based media features. MOWL supports probabilistic reasoning with Bayesian networks in contrast to crisp Description Logic-based reasoning with traditional ontology languages. We compute the joint probability distributions of the concept and the media nodes and apply the FBN technique to create the probabilistic associations. The technique is applied periodically as newly labelled multimedia data instances are added to the collection and the ontology is updated. This semi-automated maintenance of ontology alleviates significant efforts on the part of knowledge engineers.
3. **Application** The multimedia ontology is used for *a*nnotation generation for new instances of digital artefacts. A set of media feature classifiers are used to detect the media patterns corresponding to the media nodes in the ontology. The MOWL ontology can then be used to recognize the abstract domain concepts using a probabilistic reasoning framework. The concepts so recognized are used to annotate the multimedia artefacts. The goal behind building such a framework is to give a novel multimedia experience to the user seeking to retrieve resources belonging to a digital collection. The conceptual annotations are used to create semantic hyper-links in the digital collection, which along with the multimedia ontology, provide an effective *s*emantic browsing interface to the user.
4. **Evaluation** As the multimedia ontology is created and maintained along with the building of the digital collection, each process in our framework is constantly evaluated for integrity and scalability. Users and domain experts are part of the process of updating the knowledge base as new learning takes place and changes happen in the real world.

5 Ontology learning from multimedia data

Ontology learning not only improves the efficiency of ontology development process, but also enables discovery of new knowledge by tapping into data repositories. The data available with multimedia collections are of two kinds: *T*extual meta-data which gives additional knowledge about the content, and *C*ontent-based features extracted from the multimedia data. There are different approaches to multimedia data handling using either or both these kinds of data. Success in multimedia analysis and retrieval applications has been seen to occur in those methodologies which effectively combine these to complement each other. In this section, we describe how we are able to use both kinds of data to refine the basic multimedia ontology constructed at the previous stage.

There are two inputs to the Ontology Learning process—a basic ontology of the domain which is constructed with the help of knowledge provided by the domain experts, and conceptual annotations by domain experts, based on observable parameters in the media files. We illustrate the ontology learning by taking a simple example snippet from the basic ICD ontology Γ_B shown in Fig. 3. This seed ontology for the ICD domain is initially constructed by encoding specialized knowledge gathered from the domain experts. Next step involves annotating the training data and correlating media features. This snippet, enriched with media features (pink elliptical nodes), shows `BharatnatayamDance` and `OdissiDance` as two styles of `IndianClassicalDance`. `BharatnatyamDance` is related to the music form `CarnaticMusic` and a concept `Teermanam` which is a dance step typically contained in `BharatnatyamDance` performances. Media manifestations of `CarnaticMusic` include a musical beat called `AdiTaal` and an instrument `MridangamInstrument` which is regularly played as part of a Carnatic music performance. The concepts related to `OdissiDance` are the music accompanying its performances which is `OriyaMusic`, and the concept `Chawk` which has media manifestations in the form of a posture `ChawkPosture` and a dance step `ChawkBhramariDanceStep`. MOWL encoding of the ontology is done to associate the expected media patterns with concepts as well as to associate probability values to the CPTs. Some of the probability values come from the domain experts' perspective, while the others are obtained from the training set of videos. The pair of values at each link in the ontology denote the conditional probabilities $P(M \mid C)$ and $P(M \mid \neg C)$, where C is a concept and M represents an associated concept or media pattern.

Uncertainty specification is supported in MOWL, and Bayesian network reasoning is possible with observation models derived out of a MOWL ontology. With this fact in mind, we explain the learning of our MOWL ontology in terms of Bayesian network learning. We apply a standard BN learning algorithm and extend it to learn uncertainty between concepts and their media properties. The basic structure of the BN for a concept, which is the start point of the learning, comes from its OM drawn from the basic domain ontology in MOWL. The BN is learnt using the training set of annotated videos which provide the case data for learning. The learnt BN may have some new links between the nodes while some older links may be deleted if the causal dependency between the two nodes is below a threshold. Once the BNs are learnt, the learning is then applied to update the structure and uncertainties encoded in the basic MOWL ontology. The efficacy of learning is tested by building applications of annotating, searching and browsing based on the learnt ontology and testing for expected improvement in results.

5.1 Bayesian network learning

A Bayesian network is characterized by its topology and the conditional probability tables (CPTs) associated with its nodes. The goal of learning a BN is to determine both the structure of the network (structure learning) and the set of CPTs (parameter learning). An OM in our scheme, modelled as a Bayesian network, is in effect, a specification for the concept in terms of searchable media patterns. The joint probability distribution tables that signify causal strength (significance) of the different media properties towards recognizing the concept are computed from the probabilistic associations specified in the ontology. Thus, we have already got a basic structure of the BN with CPTs reflecting the domain experts' knowledge of the domain. Our aim is to use the learning algorithm—to refine this structure, which includes (1) discovering new links or relationships between concepts, and (2) removing some obsolete links, i.e. getting rid of some properties or relationships which do not exist in data; and—to learn the parameters of the Bayesian network. The algorithm must take into account the media observables or features that are associated with the concept nodes.

For our learning scheme, we have selected a standard Bayesian learning technique called the Full Bayesian Network learning [20]. We have extended this algorithm to learn structure and parameters of the Bayesian networks which correspond to the OMs for concepts in our multimedia ontology. The data for learning come from the training set of videos using the media-based features of the examples which help assign values to the variables in the network.

BN structure learning often has high computational complexity, since the number of possible structures is extremely huge. FBN overcomes the bottleneck of structure learning by not using the structure to represent variable independence. Instead, all variables are assumed dependent and a full BN is used as the structure of the target BN. FBN learning uses decision trees as the representation of CPTs [6] to obtain a more compact representation. The decision trees in CPTs are called CPT-trees. In learning an FBN classifier, learning the CPT-trees captures essentially both variable independence and context-specific independence (CSI).

For each OM extracted from the base ontology Γ_B, a set of subnets, each of which is a naive Bayesian network and has a height = 1, is obtained. The CPTs are copied into each subnet from the OM. FBN learning from case data takes place in each subnet, updating the structure and the parameters of the BN. The learnt subnets then update the OM and the learnt OMs are used to update the ontology. Figure 4a shows a subnet which is a naive Bayesian network, constructed from a snippet of the MOWL ontology, showing a concept node C, related to some other concepts X_i by MOWL relations. X_i are further connected to some media nodes shown as leaf nodes F_is in the snippet. These denote

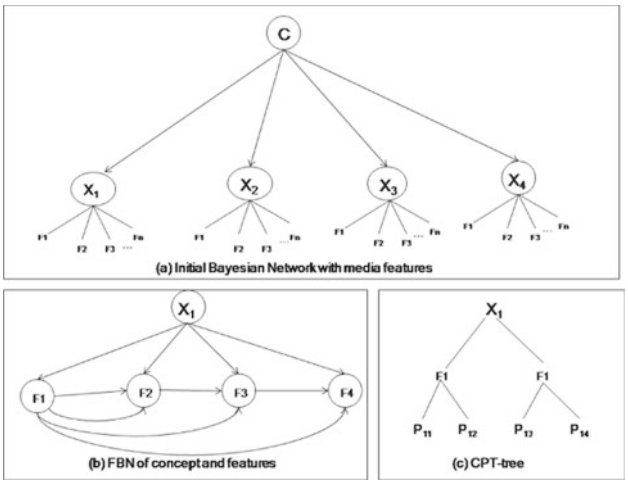

Fig. 4 Full Bayesian network with observable media features

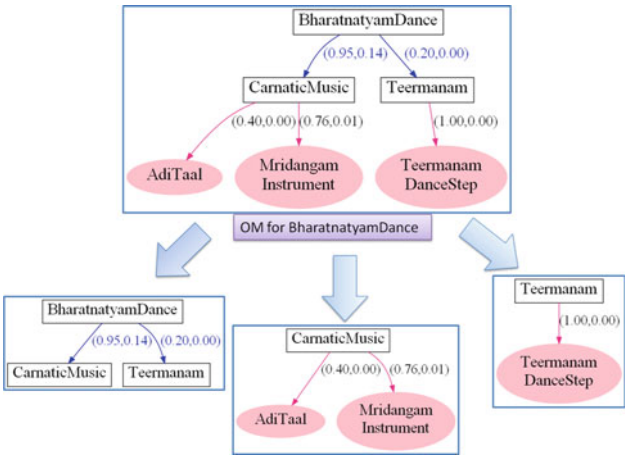

Fig. 5 Observation Model of concept `BharatnatyamDance` from ICD ontology Γ_B, split into its subnets for FBN learning

the media observable features associated with the concept. For example, the OM for concept node `BharatnatyamDance` (Fig. 5), has `Carnatic Music` as a related concept, which is further associated with the observation of the media patterns `AdiTaal` and `MridangamInstrument`. In this figure, we show how the OM for `BharatnataymDance` is split into naive Bayesian subnets. FBN structure learning in Sect. 5.2 explains the learning of each subnet, in terms of learning an FBN classifier, for a concept C and attributes X_i, with CPT learning. In Sect. 5.4, we have extended the FBN learning algorithm to that part of the BN, where observation of media features is associated with high-level concepts.

5.2 FBN structure learning

Given a training data set S, we partition S into |C| subsets, each S_c of which corresponds to the concept value c, and then construct an FBN for S_c. Learning the structure of a full BN actually means learning an order of variables and then adding arcs from a variable to all the variables ranked after it. The order of the variables is learnt based on total influence of each variable on other variables. The influence (dependency) between two variables can be measured by mutual information defined as follows:

$$M(X, Y) = \Sigma_{xy} P(X, Y) \log P(X, Y) \quad (1)$$

where x and y are the values of variables X and Y, respectively. Since we compute the mutual information in each subset S_c of the training set, $M(X, Y)$ is actually the conditional mutual information M(X,Y|c). This ensures a high probability of learning true dependencies between variables. In practice, it is possible that the dependency between two variables, measured by Eq. 1, is caused by noise. Thus, a threshold value is required to judge if the dependency between two variables is reliable. One typical approach to defining the threshold is based on the Minimum Description Length (MDL) principle. Friedman and Yakhini [7] show that the average cross entropy error is asymptotically proportional to $log N/2N$ where N is the size of the training data. Their strategy is adopted to define the threshold to filter out unreliable dependencies as follows:

$$\varphi(X, Y) = \log 2N/2N * T_{ij} \quad (2)$$

where $T_{ij} = |X_i| \times |X_j|$, $|X_i|$ is the number of possible values of X_i, and $|X_j|$ is the number of possible values of X_j. In structure learning algorithm the dependency between X_i and X_j is taken into account only if $M(X_i; X_j) > \varphi(X_i, X_j)$. The total influence of a variable X_i on all other variables denoted by $W(X_i)$ defined as follows:

$$W(X_i) = \sum_{j(j \neq i)}^{M(X_i;X_j) > Q(X_i,X_j)} M(X_i; X_j) \quad (3)$$

Equation 3 for concepts in `BharatnatyamDance` subnet in Fig. 6, computes W(`BharatnatyamDance`) > W(`CarnaticMusic`) > W(`Teermanam`). Accordingly, an ordering is imposed on the nodes in the subnet, to generate a new structure. Once CPTs are learnt, as detailed in the next section, the parameters determine whether all the links are retained, or some are deleted. Figures 5, 6 and 7 show the splitting of the OM for concept `BharatnatyamDance`, its FBN learning and updation afterwards.

5.3 Learning CPT-trees

After the structure of an FBN is determined, a CPT tree should be learned for each variable X_i. As per FBN learning, we have used the Fast decision tree learning algorithm for learning CPTs. Before the tree growing process, all the variables X_j in $\pi(X_i)$(parent set of X_j) are sorted in terms of

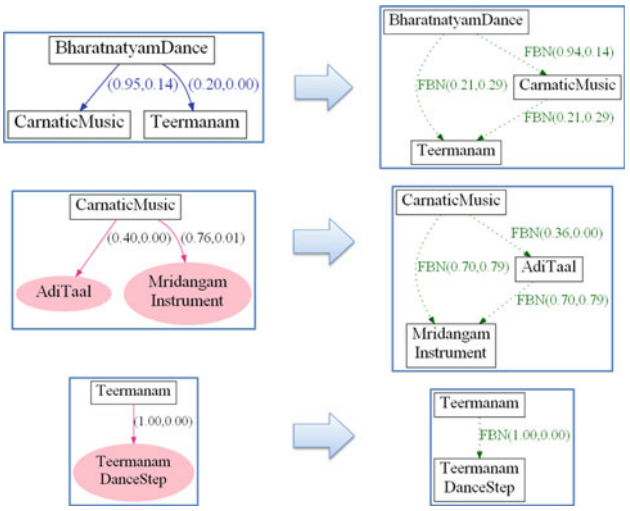

Fig. 6 Subnets of `BharatnatyamDance` OM updated with FBN learning.

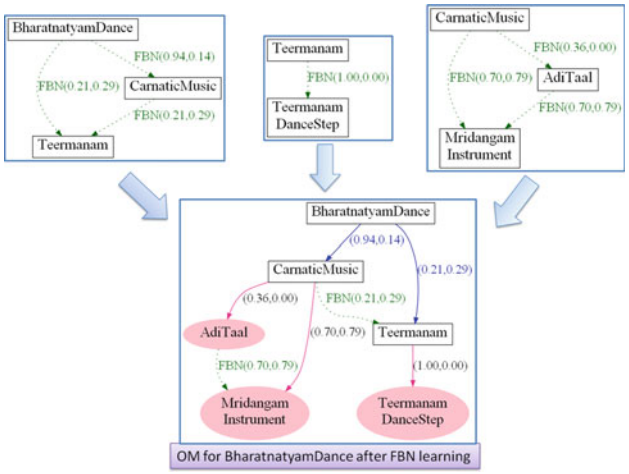

Fig. 7 Observation Model of concept `BharatnatyamDance` updated after FBN learning

mutual information $\mathrm{M}(X_i, X_j)$ on the whole training data, which determines a fixed order of variables. In the tree growing process, the variable X_j with the highest mutual information is removed from $\pi(X_i)$, and the local mutual information $M^S(X_i, X_j)$ on the current training data S, is computed. If it is greater than the local threshold $\varphi^S(X_i, X_j)$, X_j is used as the root, and the current training data S is partitioned according to the values of X_j and this process is repeated for each branch (subset). The key point here is that, for each variable, the local mutual information and local threshold is computed only once. Whether failed or not, it is removed from X_i and is never considered again. The fast CPT-tree learning algorithm can also be found in [20].

5.4 Learning associations of observables with concepts

We have extended the FBN learning algorithm to learn associations of concepts with observables features. Figure 4b shows a concept node X1 with associated media properties F1–F4 as its children. An FBN is constructed for each value x_i of $X1$ denoting an ordering amongst media features. CPT-trees denoting uncertainties between a concept and its media properties are learnt using the same algorithm as for learning uncertainties between concepts. The basis of the FBN algorithm is the mutual information which denotes the influence (dependency) between two attributes, i.e. two media features here. This is computed by equation 4.

$$M(F_i, F_j) = \Sigma_{f_i, f_j} P(f_i, f_j) \log P(f_i, f_j) \tag{4}$$

where f_i and f_j are values for F_i and F_j, respectively. $M(F_i, F_j)$ is actually the conditional mutual information $\mathrm{M}(F_i, F_j)|x_i)$, i.e. dependency between the two features, given a value x_i of the concept X_1. To compute $P(f_i, f_j)$, we need to map the features extracted to a fixed set of symbolic features values in the features space. To recognize symbolic feature states in each feature space, we apply the following clustering scheme.

We pick a set of N randomly selected videos from our video database for clustering. Every video is randomly sub-sampled to get S samples. These samples could be single frames or a group of frames (GoFs), each consisting of a group of c continuous frames—the value of c depends on video size. Different low-level media features as required by the media classifiers are extracted for each GoF. These feature values are then clustered using *K-Means clustering* to form K clusters in each feature space. Therefore, for each feature space $N \times S$ feature values are found that are clustered to get K clusters. The K cluster center values represent K symbolic feature states or media feature 'terms' which are available in the data set. Each other video in the collection is similarly preprocessed and sub-sampled to extract media features for S GoFs in the video. By performing feature-specific similarity computations of feature values with feature 'terms', we can recognize the occurrence of these 'terms' in a video. This media-feature extraction, clustering and modelling scheme is explained in detail in [13]. Thus computation of probability $P(f_i, f_j)$ is mapped to computation of $P(c_k, d_l)$ where c_k and d_l denote cluster center values which f_i and f_j map to, in their respective feature spaces.

After computing mutual information between features, the FBN algorithms for structure and parameter learning can be applied, to learn the association of the concept with each feature as well as dependencies between features. The `CarnaticMusic` subnet in Fig. 6, illustrates the association of the concept `CarnaticMusic` with its media manifestations `AdiTaal` and `MridangamInstrument`, along with the new ordering amongst the media nodes, learnt through the FBN technique.

6 Application of ontology learning in a heritage domain

In this section, we illustrate the application of our ontology learning scheme in building the ontology of the ICD domain detailing each step in the construction process. We show how the ontology is constructed from domain knowledge, then fine-tuned with the help of FBN learning using labelled ICD videos. We then validate the performance of our ontology learning scheme with experiments that measure the similarity in structure between the FBN learnt ontology and an *expected* ontology as provided by the experts. Another set of experiments which validate the parametric learning have been done to *recognize* the various abstract domain concepts with the help of the learnt ontology.

6.1 'NrityaKosha' compilation

The ICD heritage collection called 'NrityaKosha' was compiled by gathering dance videos from different sources. These include a highly specialized collection called 'Symphony Celestial Gold Edition' purchased from Invis Multimedia,[2] which contains videos of classical dance performances by eminent Indian artists. Another set of high-quality dance performance videos was obtained from the Doordarshan Archives of India.[3] Many dance DVDs were donated for research purposes by reputed artists of ICD [12]. The videos contain dance and music performances, training and tutorials on different dance forms, as well as many interviews and talks on ICD. We started work with a data set of approximately 20 h of dance videos. These consist of dance performances of mainly two Indian classical dance forms—Bharatnatyam and Odissi. The ICD ontology was constructed by encoding specialized knowledge gathered from the domain experts, as well as from dance manuals like *Natya Shastra* and *Abhinaya Darpan*.

The ICD dance videos, talks and interviews provided us with additional domain knowledge to formulate this basic ontology. The ontology is written in MOWL. The experts gave their observations on a set of about 30 % ICD videos, specifying dance forms, music forms, dance postures, dance steps, hand-gestures, name of a dancer, musicians, etc., that were part of a dance performance. Other meta-data about the video snippets was collected from the DVD covers, background commentary, scrolling text (ticker) and web. A video annotation tool, which allows conceptual annotation of different parts of a video at multiple levels of granularity, was used for this purpose. It creates video annotation files in an XML format, in tune with MPEG-7-based description scheme. These were then used as a training set to fine-tune the ICD ontology by applying FBN learning. Our ICD ontology contains around 500 concepts related to Indian dance and music in the ontology, out of which about 260 have media-observable patterns (features/examples) associated with them. Based on the expert observations, video frames showing dance postures, short video clips containing dance actions, wav files for music forms, etc., were extracted from the training set of videos. These multimedia files were attached as multimedia examples to the relevant domain concepts in the ICD ontology, and were also used as training data to train media detectors.

The ontology learning which happens in our framework has two aspects:

- learning the structure of the ontology, which involves addition and deletion of links in the ontology, thus changing the causal dependencies between concepts, and between concepts and media nodes.
- learning the parameters which are the conditional probabilities of the causal relations in the ontology.

In this section, we illustrate the validation of learning the *structure* and the *parameters* of a multimedia ontology, using ICD ontology as an example. The *parameters* are simultaneously learnt in the FBN algorithm, and are verified with demonstration of concept-recognition and semantic annotation generated as its consequence.

6.2 Learning the structure of ICD ontology

To conduct the experiments for validating the learning of ontology, we need an *expected* version of the ontology as a benchmark, with which we can compare the learnt ontology and verify that the structure learnt is valid given a bounded error margin. The starting point in the ontology learning process, is the basic ontology Γ_B, constructed from domain knowledge obtained from dance gurus (teachers and masters), and enriched with multimedia data from the labelled examples. This ontology represents the domain experts' perspective and encodes the complexities of the background knowledge of the heritage ICD domain.

We obtained an expected version of the ontology Γ_E shown in Fig. 8 from a different set of domain experts—the Indian classical dancers who have contributed their dance videos to the ICD heritage collection. Their version of the ontology differs in structure with Γ_B, as the domain concepts and their relationships, as interpreted by the dancers are more in tune with the current context in which the dance performances take place. Indian classical dance domain being a heritage domain, the dancers do not have the liberty of adapting the rules and grammar of the domain beyond a certain permitted boundary, but they do understand the practical dependencies and corelations between dance, music, pos-

[2] http://www.invismultimedia.com.

[3] http://www.ddindia.gov.in/About%20DD/Programme%20Archives.

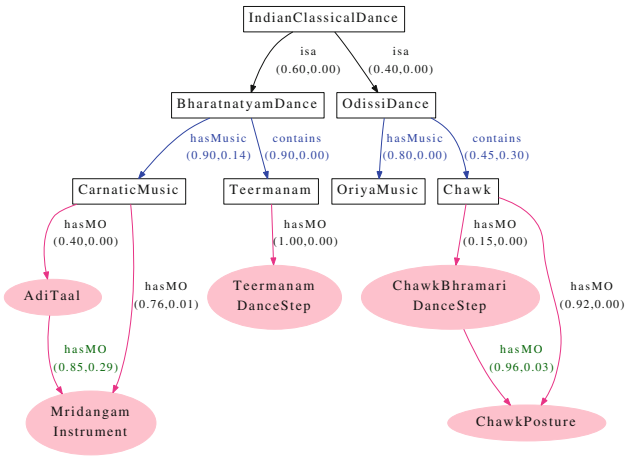

Fig. 8 Expected ontology snippet Γ_E of the ICD domain as specified by the Indian classical dancers

tures, themes and roles in the existing scenario *better* than the theoretical knowledge that the dance gurus might posses.

We perform the FBN learning on observation models obtained from Γ_B, then apply that learning to Γ_B to obtain the learnt ICD ontology Γ_L. A *graph matching* performance measure is applied to measure the similarity between the two versions of the ontology: Γ_L and Γ_E.

6.2.1 Performance measure

A MOWL ontology is a directed, labeled graph, and so are the two versions of the ontology—Γ_L and Γ_E, which need to be measured for similarity. There are several standard similarity measures defined to compute similarity between directed, labeled graphs, which are graphs with a finite number of nodes, or vertices, and a finite number of directed edges. We have chosen *graph edit distance* and *maximum common sub-graph* reviewed in [3]. A maximum common sub-graph of two graphs, g and g', is a graph g'' that is a sub-graph of both g and g' and with the maximum number of nodes, from among all possible sub-graphs of g and g'. The maximum common sub-graph of two graphs need not be unique. The larger the number of nodes in the *maximum common sub-graph* of two graphs, the greater is their similarity.

The other performance measure *graph edit distance* provides more error-tolerant graph matching. A graph edit operation is typically a deletion, insertion, or substitution (i.e. label change), and can be applied to nodes as well as to edges. The edit distance of two graphs, g and g', is defined as the "shortest sequence of edit operations" that transform g into g'. Obviously, the shorter this sequence is, more similar are the two graphs. Thus, edit distance is suitable to measure the similarity of graphs. According to [3], the *maximum common sub-graph* g'' of two graphs g and g' and their *edit distance* are related to each other through the simpleequation

$$d(g, g') = |g| + |g'| - 2|g''| \tag{5}$$

where $|g|$, $|g'|$ and $|g''|$ denote the number of nodes of g, g' and g'', respectively.

6.2.2 Logic and implementation

The process of applying ontology learning in terms of obtaining the OMs from ontology, learning the OMs and then updating the ontology with the changed structure and parameters is detailed in algorithm 1 The two inputs to this algorithm are the basic ontology Γ_B and case data obtained from the labelled set of files from the multimedia collection, in our case the labelled videos from the ICD collection. As mentioned in Sect. 5.1, a set of naive Bayesian subnets, each of height = 1, is obtained for each OM extracted from Γ_B. The CPTs are copied into each subnet from Γ_B. FBN learning from case data takes place in each subnet, updating the structure and the parameters of the BN. The learnt subnets then update the OM and the learnt OMs are used to update the ontology. The output of the algorithm is the learnt ontology Γ_L.

Algorithm 1 Applying FBN learning to learn an ontology.

Require: a) Basic Ontology Γ_B
b) Case Data obtained from the training Set of Multimedia Documents
Ensure: Learnt Ontology Γ_L

```
1: procedure MAIN
2:     Compute 𝒪 = { Ω : Ω is an OM for a concept in Γ_B}
3:     for i = 1 to |𝒪| do
4:         Compute 𝒮 = { s : s is a subnet of height = 1 in Ω_i }
5:         for j = 1 to |𝒮| do
6:             Apply FBN learning with case data to learn s_j
7:             Update Ω_i with s_j
8:         end for
9:     end for
10:    Γ_L = Γ_B
11:    for i = 1 to |𝒪| do
12:        Update Γ_L with Ω_i
13:    end for
14:    Return Γ_L
15: end procedure
```

The implementation of the Bayesian network learning applied here, was done with the Netica Java API (NeticaJ).[4] Figure 8 shows the *expected* version of ICD ontology. The main difference with the specifications in Γ_B are as follows:

- `Chawk Bhramari dance step` contains the `Chawk posture`.
- `Adi taal` musical beat has `Mridangam instrument` as its media observable, as the latter is often used in Bharatnatyam dance performances to play the former. Some of the probabilities specified in the two ontologies are also

[4] http://www.norsys.com/netica-j.html.

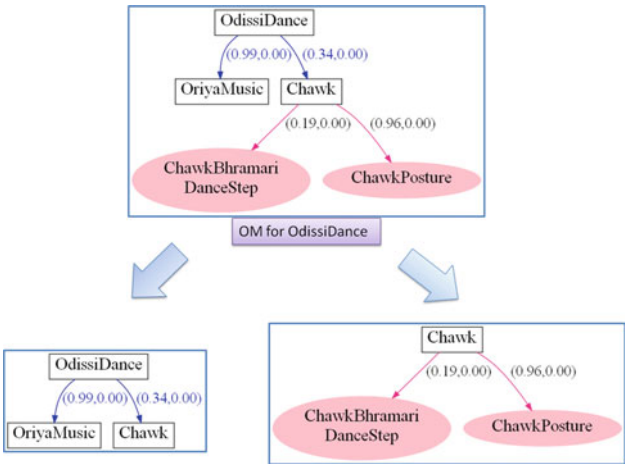

Fig. 9 Observation Model of concept `OdissiDance` from Γ_B, split into its subnets for FBN learning

different, but we are not looking at parametric similarity here.

Similar to the learning of `BharatnatyamDance` subnet in Sect. 5.2, here we show the splitting of the OM for `OdissiDance`, its FBN learning and updation in Figs. 9 and 10. Figure 11 shows the learnt ICD ontology Γ_L which is constructed after the updated OMs of concepts *Bharatnatyam* and *odissidance* have been used to update the basic ontology Γ_B.

Applying the graph similarity performance measures, we first find the *maximum common subgraph* Γ_C of the two graphs Γ_E and Γ_L. Then *graph edit distance* between the two graphs is computed as follows:

$$d(\Gamma_E, \Gamma_L) = |\Gamma_E| + |\Gamma_L| - 2|\Gamma_C| \tag{6}$$

where $|\Gamma_E|$, $|\Gamma_L|$ and $|\Gamma_C|$ denote the number of nodes of Γ_E, Γ_L and Γ_C, respectively. As we can see here the $d(\Gamma_E, \Gamma_L) = 2$ for the ICD example snippet ontology shown here. Out of the approximately 500 concepts in the ICD ontology, there are around 182 concepts which are at a suitably high abstract level where their observation models can be tuned with the FBN learning algorithm. We experimented with 75 observation models, with number of nodes in the OMs ranging from 6 to 10 and the number of edges ranging from 5 to 10. We obtained an average performance of *graph edit distance* = 2.4 between the learnt and expected versions.

6.3 Parametric learning of ICD ontology and concept-recognition

FBN learning from case data leads to change in the structure and parameters of the Bayesian network representing the OM extracted from the ontology. After all the OMs have been learnt, they are used to update the structure of the ontology and also change the joint probabilities encoded in the ontology according to the new parameters learnt in the OMs. The ontology learnt in this manner is dynamic, as it can be refined and fine-tuned automatically with additions to the video database. The newly learnt ontology can then be applied afresh to recognize concepts in the video database. If the learning is good, then the concept-detection and subsequent annotation generation should show improved results with the fine-tuned ontology. We discuss a small example from the ICD domain to demonstrate how concept-detection improves with the domain ontology changed after applying FBN learning.

6.3.1 Concept-recognition using MOWL reasoning

Once an OM for a semantic concept is generated from a MOWL ontology, the presence of expected media patterns can be detected in a digital multimedia artefact using appropriate media detector tools. Such *observations* lead to instantiation of some of the media nodes in the OM, which in turn, result in belief propagation in the Bayesian network. The posterior probability of the concept node as a result of such belief propagation, represents the degree of belief in the presence of the semantic concept in the multimedia artefact.

For e.g., in Fig. 12, the BN corresponding to the OM of concept `Mangalacharan` is shown after some media patterns have been detected in an Odissi dance video and corresponding media nodes have been instantiated. The links between

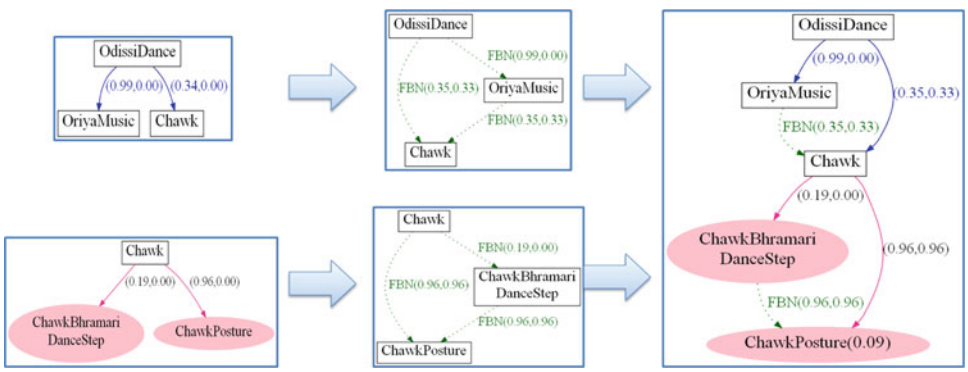

Fig. 10 Observation Model of `OdissiDance`, with its FBN learning and updation

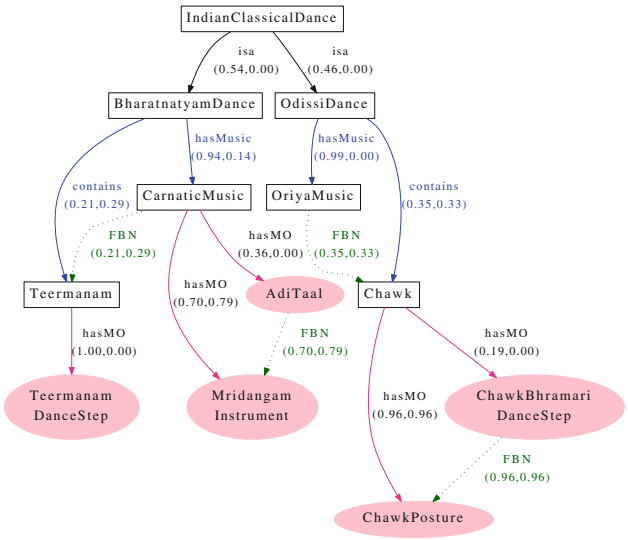

Fig. 11 Learnt ontology Γ_L of the ICD domain with its structure and parameters changed due to FBN learning from case data

concept nodes, between media nodes and between a concept and a media node denote causal relations as well as uncertainty specifications that have been learnt from data. Bracketed value with the name of each node denotes its posterior probability after media nodes have been instantiated and belief propagation has taken place in the BN. In this video, the media patterns detected with the help of concept-detectors are `ChawkPosture` and `PranamPosture`, shown as dark pink ellipses.

6.3.2 *Concept-recognition after parametric learning*

Figure 13 shows the OM of the ICD concept `Mangalacharan` after **FBN** learning has been applied to it. The OM is constructed from the ICD ontology refined with FBN learning, so the probability values shown correspond to real-world data. After applying FBN learning, some new relations (shown with *green* links and labelled *FBN*) were added, based on statistical evidence in case data.

Let the media patterns detected in the test *Odissi dance* video be `ChawkPosture` and `PranamPosture`. The corresponding media nodes are instantiated in the `Mangalacharan` OM generated from the FBN learnt ICD ontology. As in the earlier case, `Chawk` and `Pranam` are the low-level concepts which are recognized due to presence of these media patterns in data. Due to an FBN link between `PranamPosture` and `PranamDanceStep`, the latter node is also instantiated, thus leading to higher belief in the presence of concept `BhumiPranam` in the video. Higher level concept nodes

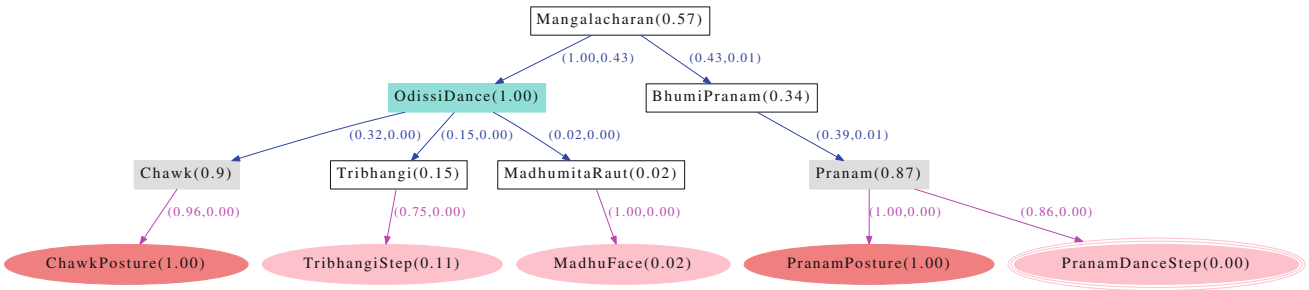

Fig. 12 Concept-recognition in the *Mangalacharan* OM generated from the basic ontology Γ_B

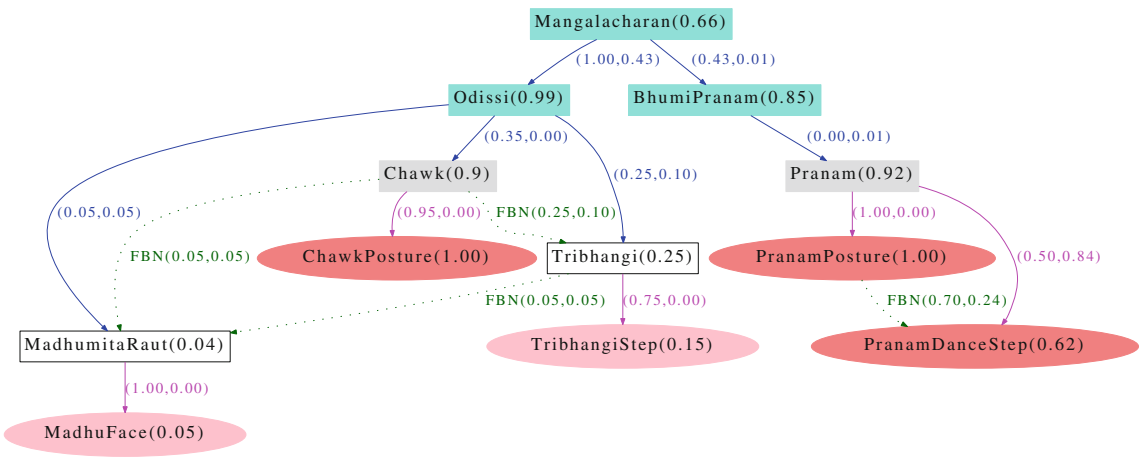

Fig. 13 Bayesian network corresponding to the observation model of the concept *Mangalacharan* after FBN Learning has caused the structure and parameters to be updated

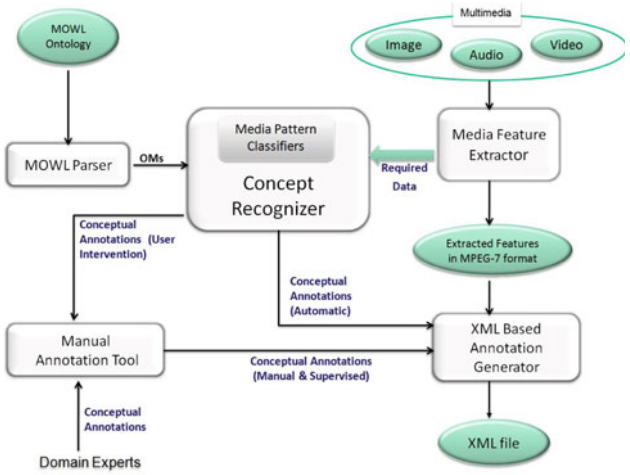

Fig. 14 Architecture of annotation generation framework

(in cyan color) are recognized to be present due to belief propagation in the BN. `Chawk` pattern causes `Odissi Dance` to be recognized. Presence of `Pranam` and `BhumiPranam` lead to recognition of `Mangalacharan` concept which is further confirmed by recognition of `Odissi Dance` concept in the video. This concept-recognition is confirmed by the labels that the domain experts had provided for the test *Odissi dance* video. Thus FBN learning has led to an improvement in concept-recognition as in the basic ontology, only one abstract concept *Odissi Dance* was recognized in the video.

6.3.3 Semantic annotation generation

An important contribution of our ontology learning is the attachment of conceptual annotation tomultimedia data belonging to a domain, thus preserving its background knowledge and enhancing the usability of this data through digital access. Figure 14 shows the architecture of our Annotation Generation framework. It consists of 5 functional components. The basis of this whole framework is the MOWL ontology created from domain knowledge, enriched with multimedia data and then refined with learning from annotated examples of the domain. The most important component of this process is the *Concept-Recognizer*. The task of this module is to recognize the high-level semantic concepts in multimedia data with the help of low-level media-based features. OMs for the high-level concepts selected by the curator of the collection, are generated from the MOWL ontology by the *MOWL Parser* and given as input to this module. Low-level media features (SIFT features, Spatio-temporal interest points, MFCC features, etc.) are extracted from the digital artefacts which can be in different formats (image, audio, video), and provided to the **Concept Recognizer** by *Media Feature Extractor*. *Media Pattern Classifiers*, trained by feature vectors extracted from the training set of multimedia data, help detect the media patterns (objects, shapes, postures, actions, music, etc.) in the digital artefacts. Some of these classifiers are detailed in our work [13]. In initial stages of building the ontology, data are labelled with the help of **manual annotations**, provided by the domain experts in XML format.

To recognize concepts in a new video of ICD, evidence is gathered at the leaf nodes, as different media features are recognized or classified in the video by the media classifiers. If evidence at a node is above a threshold, the media feature node is instantiated. These instantiations result in belief propagation in the Bayesian network, and posterior probability at

Table 1 Table to show high-level annotation results using basic and FBN learnt ontology

Concept	Basic					FBN				
	Correct	Miss	False	Precision	Recall	Correct	Miss	False	Precision	Recall
Pranam	43	10	23	0.65	0.81	55	8	13	0.80	0.87
Krishna Role	53	35	10	0.84	0.60	73	10	15	0.83	0.88
Mangalacharan Dance	23	36	13	0.64	0.39	53	14	5	0.91	0.79
Yashoda Role	7	2	1	0.87	0.77	5	3	2	0.71	0.63
Carnatic Music	92	10	3	0.97	0.90	97	5	3	0.97	0.95
Battu Dance	15	3	12	0.55	0.83	22	4	4	0.85	0.85
Adi Taal	52	12	20	0.72	0.81	54	18	12	0.82	0.75
Vanshika Chawla	22	5	3	0.88	0.81	27	2	1	0.96	0.93
Madhumita Raut	11	10	12	0.48	0.52	22	2	9	0.71	0.92
Naughty Krishna	6	3	5	0.54	0.67	5	4	5	0.50	0.55
Group Dance	12	13	9	0.57	0.48	27	3	4	0.87	0.90
Solo Dance	26	13	18	0.59	0.66	41	8	8	0.84	0.84
Krishna Sakhi Theme	1	2	3	0.25	0.33	2	1	3	0.40	0.67
Mahabharat Theme	7	15	3	0.70	0.32	15	3	7	0.68	0.83

the associated concept nodes is computed. After belief propagation, these nodes have high posterior probability. As they get instantiated, we find high belief for the existence of other high-level abstract nodes. Conceptual annotations are generated and attached to the video through Semantic annotation generation. Results for some of the conceptual annotations generated for the ICD domain using the basic ontology and then the FBN learnt ontology are shown in Table 1. On an average, we see an improvement in *Precision* and *Recall* from the FBN learnt ontology over the results from the basic ontology.

7 Concluding remarks

In this paper, we have presented a technique by which the knowledge obtained to construct an ontology from a domain expert can be fine-tuned by applying learning from real-world examples belonging to the domain. An ontology refined in this manner is a better structured, more realistic model of the domain that it represents. In this paper, we have introduced a novel technique to populate an intangible heritage collection of videos in a scholarly domain like Indian classical dance, in order to provide a flexible, ontology-driven access to the users of the domain. The ontology learnt from video examples represents a domain more attuned to the real-world data. The Bayesian network learning technique that we have used allows us to learn new structure as well as the parameters of the Bayesian network. The ontology learnt in this manner not only has a more refined *structure* as is proved in experiments done in this paper, but also has more accurate conditional *probabilities* encoded in the CPTs attached to its concepts and media-based nodes.

References

1. Akoush S, Sameh A (2007) Mobile user movement prediction using bayesian learning for neural networks. In: Proceedings of the 2007 international conference on wireless communications and mobile computing (IWCMC '07), pp 191–196
2. Binder J, Koller D, Russell SJ, Kanazawa K (1997) Adaptive probabilistic networks with hidden variables. Mach Learn 29(2–3): 213–244
3. Horst Bunke (2000) Graph matching: theoretical foundations, algorithms, and applications. Vision Interface (VI2000), pp 082–088. http://www.cipprs.org/papers/VI/VI2000/vi2000.html
4. Buntine W (1996) A guide to the literature on learning probabilistic networks from data. IEEE Trans Knowl Data Eng 8(2):195–210
5. Ding Z, Peng Y (2004) A probabilistic extension to ontology language owl. In: Proceedings of the 37th annual Hawaii international conference on system sciences (HICSS'04), Track 4
6. Friedman N, Goldszmidt M (1996) Learning bayesian networks with local structure. In: Proceedings of the twelth conference on uncertainty in artificial intelligence, pp 252–262
7. Friedman N, Yakhin (1996) On the sample complexity of learning bayesian networks. In: Proceedings of the twelth conference on uncertainty in artificial intelligence, pp 274–282
8. Ghosh H, Chaudhury S, Kashyap K, Maiti B (2007) Ontology specification and integration for multimedia applications. Springer, Berlin
9. Heckerman D (1999) A tutorial on learning with bayesian networks. Learn Graph Models, pp 301–354
10. Hunter J, Drennan J, Little S (2004) Realizing the hydrogen economy through semantic web technologies. In: IEEE Intell Syst
11. Kojima A, Tamura T, Fukunaga K (2002) Natural language description of human activities from video images based on concept hierarchy of actions. Int J Comput Vis 50(2):171–184
12. MadhumitaRaut (2012) Madhumita raut of jayantika–the mayadhar school of odissi dance
13. Mallik A, Chaudhury S, Ghosh H (2011) Nrityakosha: preserving the intangible heritage of indian classical dance. JOCCH 4(3):11
14. Mezaris V, Kompatsiaris I, Boulgouris NV, Strintzis MG (2004) Real-time compressed-domain spatiotemporal segmentation and ontologies fro video indexing and retrieval. IEEE Trans Circ Syst Video Technol 14(5):606–620
15. Naphade M, Smith JR, Tesic J, Chang SF, Hsu W, Kennedy L, Hauptmann A, Curtis J (2006) Large-scale concept ontology for multimedia. IEEE Multimedia 13:86–91. doi:10.1109/MMUL.2006.63
16. Navigli R, Velardi P, Gangemi A (2003) Ontology learning and its application to automated terminology translation. IEEE Intell Syst 18(1):22–31
17. Neuman L, Kozlowski J, Zgrzywa A (2004) Information retrieval using bayesian networks. Comput Sci ICCS 2004:521–528
18. Niculescu S, Mitchell T, Rao R (2006) Bayesian network learning with parameter constraints. J Mach Learn Res 7:1357–1383
19. Ramachandran S, Mooney RJ (1996) Revising bayesian network parameters using backpropagation. In: Proceedings of the IEEE international conference on neural networks, pp 82–87
20. Su J, Zhang H (2006) Full bayesian network classifiers. In: Proceedings of the 23rd international conference on machine learning, pp 897–904
21. Town C (2004) Ontology-driven bayesian networks for dynamic scene understanding. In: Proceedings of the 2004 conference on computer vision and pattern recognition workshop (CVPRW'04), vol 7
22. Wattamwar SS, Ghosh H (2008) Spatio-temporal query for multimedia databases. In: Proceeding of the 2nd ACM workshop on multimedia semantics (MS '08). ACM, New York, pp 48–55. doi:10.1145/1460676.1460686
23. Zheng Z (2000) Lazy learning of bayesian rules. Mach Learn 41(1):53–84
24. Zhou L (2007) Ontology learning: state of the art and open issues. Inf Technol Manage 8(3):241–252
25. Zhou L, Booker Q, Zhang D (2002) Toward rapid ontology development for underdeveloped domains. In: Proceedings of the 35th annual Hawaii international conference on system sciences (HICSS '02), vol 4, p 106