Multi-entity bayesian networks for knowledge-driven analysis of ICH content

Giannis Chantas[1], Alexandros Kitsikidis[1], Spiros Nikolopoulos[1], Kosmas Dimitropoulos[1], Stella Douka[2], Ioannis Kompatsiaris[1], and Nikos Grammalidis[1]

[1] Centre for Research and Technology Hellas, Information Technologies Institute, 6th km Xarilaou-Thermi, Thessaloniki, Greece

[2] Department of Physical Education and Sport Science, Aristotle University of Thessaloniki, Greece

Abstract. In this paper we introduce Multi-Entity Bayesian Networks (MEBNs) as the means to combine first-order logic with probabilistic inference and facilitate the semantic analysis of Intangible Cultural Heritage (ICH) content. First, we mention the need to capture and maintain ICH manifestations for the safeguarding of cultural treasures. Second, we present the MEBN models and stress their key features that can be used as a powerful tool for the aforementioned cause. Third, we present the methodology followed to build a MEBN model for the analysis of a traditional dance. Finally, we compare the efficiency of our MEBN model with that of a simple Bayesian network and demonstrate its superiority in cases that demand for situation-specific treatment.

Keywords: semantic analysis, intangible cultural heritage, multi-entity bayesian networks

1 Introduction

By the age of six, humans recognize more than 104 semantic concepts [1] and keep learning more throughout their life. Can a computer program learn how to recognize semantic concepts in multimedia content the way a human does? In adressing this question, divergent approaches have been proposed, relying either on the use of explicit knowledge or the abundant availability of data. Advocating the former, [2], [3] are two notable cases where a small number of examples used during learning are able to provide models with sufficient generalization ability. The authors rely on the hypothesis that once a few visual categories have been learned with significant cost, some information may be abstracted from the process to make learning further categories more efficient. Taking a different perspective, the authors of [4] claim that with the availability of overwhelming amounts of data many problems can be solved without the need for complex parametric algorithms. The authors index a large dataset of 79 million images and using nearest neighbor matching for image annotation, they claim that given the excessive volume of the indexed images it is reasonable to assume that almost every "unseen" image will be close enough to a ''seen" image. These examples demonstrate the debate around the mechanism of building perceptual models and the discussion on how much of the knowledge should come in an explicit form and how much can be obtained implicitly from the available training samples. Although moving towards the one or the other extreme of the debate may still produce non-trivial recognition models, higher levels of efficiency can only be achieved if explicit and implicit knowledge are effectively combined.

The aim of this work is to verify the aforementioned statement in the specialized domain of Intangible Cultural Heritage (ICH). The term intangible cultural heritage (ICH) (UNESCO, 2013) refers to valuable traditional art forms and creative practices, such as singing, dancing, craftsmanship, etc. Preserving this knowledge is considered particularly important and the use of technology to achieve this objective has become a popular research topic. In this paper, we advocate the use of Multi-Entity Bayesian Networks (MEBNs) [5] as an efficient scheme to facilitate the analysis of ICH content, mainly due to their ability in combining first-order logic with probability theory. The remaining of this paper is organized as follows: Section 2 describes the particularities of the ICH domain and motivates the use of MEBNs. Section 4 provides some background for the constituent elements of MEBNs. In Section 5 we describe the most important characteristics of MEBNs and argue about their appropriateness to address the particularities of ICH domain. Section 6 offers some details about the methodology adopted to implement and apply MEBNs for analyzing ICH content. Finally, Section 7 explains the results of our preliminary experimental study, while Section 8 summarizes our concluding remarks.

2 Semantic analysis in the ICH domain

The semantic analysis of digital heritage resources is considered a particularly important prerequisite for their preservation. This is even more evident in the domain of ICH. Indeed, given that during the preservation of intangible heritage the significance of heritage artifacts is implied in their context, the scope of digital preservation extends to the preservation of the background knowledge that puts them in proper perspective. For example, Mangalacharan [6] is an invocation dance in Indian Odissi dance form, which is specified in terms of specific and predefined dance actions and it is accompanied by a specific kind of music. The dance actions entail the movement of human body parts and interaction with object and the accompanying music has features that fit to the dance. Moreover, high-level concepts are manifested in the dance that are composed of basic body actions, which are related to the music features. Tsamiko dance is another example of traditional dance that shares some common features (e.g, a predefined sequence basic body actions) with Mangalacharan. Thus, generalizing by these examples, the preservation of heritage resources requires a solution to the problem of: (a) recognizing media patterns that correspond to elementary domain concepts like objects, postures, actions, audio tempos, etc, and (b) consider these elementary domain concepts as evidence that act in favor or against a hypothesis stating that the analyzed media item manifests a certain high-level concept.

The knowledge regarding the ICH domain is diverse, vast and intricate. A major difficulty in representing such knowledge is the inherent ambiguity and uncertainty in concepts prevalent in this domain. In meeting this challenge, explicitly provided logic-based rules need to be combined with a probabilistic inference framework in order to map low-level multimedia features to high-level concepts. Initially, the domain concepts and their relations will have to be expressed in a machine understandable format that should be also capable of encoding different snapshots of the analysis environment (e.g., number of dance steps). Then, low-level multimedia features that may incorporate visual, or other types of signals will have to be analyzed to obtain elementary conceptual information, acting as evidence. Finally, the framework used for probabilistic inference should inherit the logic-based rules encoded in the first step and evaluate the extracted evidence in the context of the domain knowledge. Thus, at the core of semantic multimedia analysis lies the development of a theory that will not only manage to effectively combine logic-based rules with probabilistic inference, but will also offer the necessary flexibility to cope with an un-predictable and dynamically changing environment.

3 Related work

A number of works have been presented in the literature that aim to represent knowledge in a probabilistic manner. OOBN models [7] have been proposed as an alternative to standard BN for overcoming the inherent inflexible structure of BN. An OOBN object is a collection of domain attributes that extents regular BN nodes, so as to become more flexible to situations that require customization. Probabilistic relational models (PRMs) [8] extend Bayesian networks by introducing the concept of properties, and relations between them. Like MEBNs, PRMs provides a similar mechanism to built situation specific probabilistic models. However, OOBN and PRM expressivity is inferior to MEBN, mainly due to the context limitations used to enforce logical constraints on the model variables.

Ontologies with probabilistic extensions have been also used for the semantic analysis of ICH content. For example, in [6] the authors propose the use of ontology-based mapping for linking cultural heritage content to ICH concepts. More specifically, the ontology used in this framework includes the descriptions of domain concepts that are formally given in terms of the related low-level audio-visual features, appearing in the multimedia content. In this way, a convenient semantic interpretation of the multimedia data is enabled. In another closely related work [9], a semi-automatic ontology construction methodology is proposed for combining bayesian networks with probabilistic inference. The goal of this work is to facilitate the semantic analysis of cultural Indian dances, i.e. detection of specific dance styles and moves in multimedia with cultural content. Note however, that although the ontology is constructed using probabilistic methods (i.e. as a BN of concepts and relations), the BN remains unchanged.

This is a serious modeling shortcoming that motivates the use of MEBNs. In the following, we provide the necessary background so as to advocate the use of MEBNs as a potential solution to the problem of semantic analysis in the domain of ICH.

4 Background

First-Order Logic [10] is a formal system that is used as a rigorous foundation of knowledge representation schemes. A theory in first-order logic consists of the axioms, expressed as sentences in the FOL language, in conjunction with the sentences that are derived from the axioms according to the reasoning rules, i.e., valid sentences. The main components of the FOL language are variables, functions, predicates and rules. A set containing all instantiated components of a FOL theory is called an interpretation. However, FOL does not provide expressivity to model uncertain knowledge, which is a consequence of the fact that each interpretation mentioned above shares equal validity with the others. As we will see in Section 5, the key feature of MEBN is the assignment of a probability to every interpretation.

Ontologies are a powerful tool able to express knowledge in different levels of granularity [11]. The knowledge about a domain can be expressed by a structure that associates the domain concepts and defines relations using the allowable operators, such as "Description Logics" (DL) [12]. DL constitutes a set of constructors (e.g. intersection, union, disjoint, complement, etc) that offer the expressivity to define complex knowledge of a domain, however they lack the ability to express knowledge with probabilistic terms. This has motivated the development of probabilistic ontologies as a means to encode domain knowledge and introduce uncertainty in ontology reasoning. For instance, in [13] the OWL language is augmented to allow additional probabilistic markups and a set of structural translation rules is used to convert an OWL ontology into a directed acyclic graph of a BN. Probabilistic rules are used to cope with uncertainty and ontologies combined with belief networks are employed to express and migrate into a computationally enabled framework, the semantics originating from the domain. Although dynamic, the inference potential supported by probabilistic ontologies is still restricted by the encoded snapshot of the domain knowledge, offering very limited flexibility in adapting to the situation at hand.

Bayesian Networks (BN) are stochastic models that have been applied successfully to problems where uncertainty is prevalent. BN are directed, acyclic graphical (DAG) models [14] that consists of random variables, represented by nodes, along with their relations determining the structure of the DAG. A conditional probability distribution is defined for each variable. Inference in BN refers to the process of estimating the posterior distribution for a subset of the random variables, given the observed values of another subset. The observations are direct evidence that we obtain from the real world that we are trying to model. Prior to inference there is also a training phase where the conditional probability distributions of the BN are estimated. Typically, these distributions are estimated using training data and a suitable training algorithm (e.g., maximum likelihood). The most serious drawback of BNs is that their structure is kept fixed, once they are designed. They are insufficient to model problems in dynamically changing environment since they lack the expressive power to represent entity types that can be instantiated as many times as required for the situation at hand.

5 Multi-entity Bayesian Networks

MEBN logic is a formal system that unifies probability theory and classical first-order logic (FOL). Thus, MEBNs are the outcome of the combination of BN with FOL. From a Bayesian perspective, MEBNs are extended BNs by incorporating FOL. Their main advantage is in combining the capability of BN to model uncertainty with the expressivity of FOL in representing knowledge. The key feature of MEBNs is the ability to build situation specific BNs (SSBN) that are customized according to the snapshot of the environment being modeled in an arbitrary situation. In this way, MEBNs overcome the inflexibilty of BNs to adapt to the volatile environment being model, since they have a fixed structure and conditional probability for each node.

Technically, a MEBN is a collection of MEBN fragments (MFrags). An MFrag includes (among others) *resident* node(s), for each of which a local conditional distribution and a set of parent nodes (if any) are defined. The MFrag of a resident node is called its *home* MFrag. Also, in an MFrag, there are input nodes that are resident nodes in other MFrags. The parents of a resident node can be either resident, input nodes or both. The resident nodes are, in a sense, templates that are used to construct the nodes of the SSBN, i.e., the name of the nodes, the dependencies with other nodes and the conditional probability distribution. The local conditional distribution of a resident node in an MFrag is a function that produces the conditional probabilities of the SSBN nodes produced by the resident node. This function takes as input the structure of SSBN and produces a conditional probability for the related node accordingly.

Another component of an MFrag is the logical variables, placed as arguments in resident and input nodes, and logical constraint nodes, imposing constraints on the logical variables participating in the MFrag. Logical variables and their constraints are the manifestation of the FOL into MEBN modeling. The structure of an SSBN is determined by the logical variables and the admissible by the constraints values to which they can be instantiated, according to the situation of the environment being modeled (e.g., number of nodes a resident node, acting as a template, replicate). In other words, logical variables and their constraints drive the construction of the SSBNs based on the evidence collected by the environment, translated as potential values of the logical variables.

The ultimate goal of modeling with MEBN is inference, which provides us with the ability to analyze the environment being modeled (e.g., a stochastic process or a static snapshot of a closed system), based on evidence. Inference is performed on the SSBN, which results based on evidence. There are two steps in SSBN construction. In the first step, the logical variables are instantiated to values determined by the environment and according to logical constraints. The resulting nodes in the SSBN have a defined set of parent nodes and a conditional distribution. In the second step, a subset of the SSBN random variables (i.e. the *observed* variables) are considered known (observed) and instantiated to their observed (measured) values. Then, Bayesian inference provides the posterior probability distribution of the unknown random variables we want to estimate.

6 Applying MEBN theory in the domain of ICH

In order to validate our assumptions in a real world problem, we have used MEBNs as a knowledge representation and analysis tool for recognizing the different styles of a traditional greek dance. There are two main reasons motivating the use of MEBNs for this specific task, namely, uncertainty modeling and situation-specific analysis. Uncertainty in this case is manifested in two cases. In the first, a dancer may unexpectedly deviate from the dance pattern (e.g., skips a dance step). In the second, the step detector may fail to detect a step and/or correctly recognize its features. MEBNs are capable to model both the volatility of the step number and the uncertainty (randomness) aspects of each performance. Also, the situation specific analysis capability is useful due to the fact that, usually, the number of steps is not *a priori* known. SSBN can be proven very beneficial for the dynamic modeling of such situations. In our work, the role of the MEBN is to adapt in each performance and model in a probabilistic Bayesian framework the uncertainty aspects of the dance. Based on that, the ultimate goal is to detect the dance style through probabilistic inference.

The first step in employing MEBNs is to consult the experts in order to elicit and formally encode the domain knowledge. Thus, a methodology for the ontology specification and engineering will have to be employed. Subsequently, the knowledge encoded in this ontology will act as the basis for constructing the corresponding MFrags. Then, the observations extracted from the analysis of sensor signals will be injected to the framework so as to generate the SSBN and perform probabilistic inference. Finally, decisions about the different dance styles can be made based on the posterior probability distribution of the network.

6.1 Ontology specification & engineering

Most state-of-the-art methodologies for ontology engineering incorporate the requirements specification activity. The communication tool that is used during this activity is a set of competency questions (CQs) that are posed to the experts. The CQs are answered by the experts in natural language. The terms used in these answers are subsequently analyzed with respect to their frequency and semantic affinity, so as to extract the terminology (names, adjectives and verbs) that will be formally represented in the ontology by means of concepts, attributes and relations. In accordance with this practice, we have followed the methodology of [15] in order to specify and engineer the ontology presented in Figure 1.

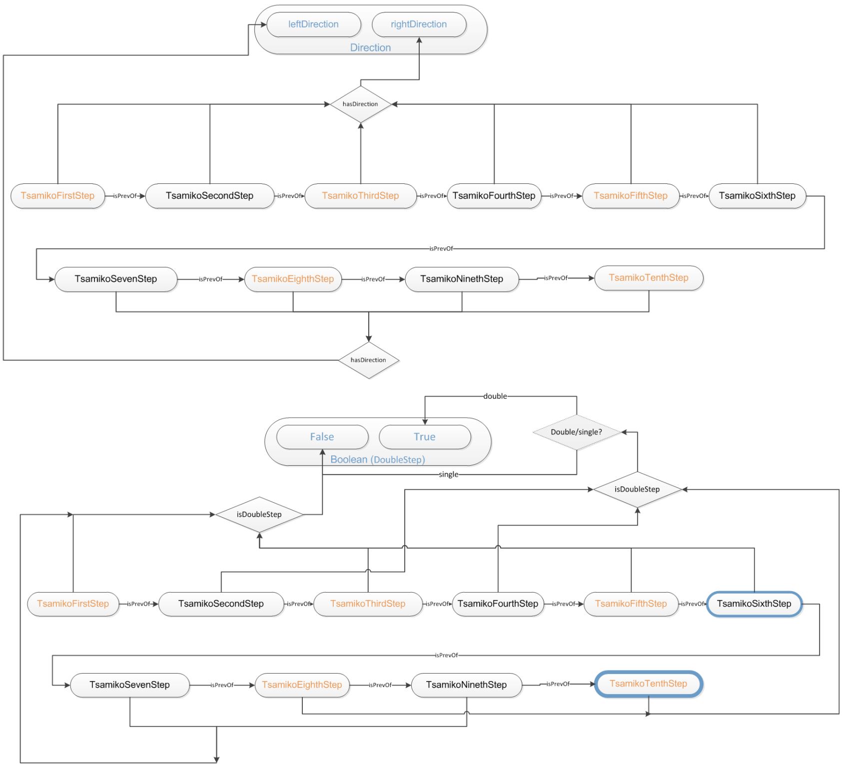

Fig. 1. Tsamiko ontology graph. Some nodes were colored for brevity of demonstration, black: right foot steps, orange: left foot steps. The TsamikoSixStep and TsamikoTenStep are high-foot steps when the dance style is male. Also, grey areas illustrate the 'hasDouble(Right/Left)Step' relation.

The style of Tsamiko dance is characterized as "double" or "single" and as "male" or "female". Thus, there are four different characterizations, and, hence, Tsamiko dance styles: (single, male), (double, male), (single, female) and (double, female). In distinguishing between the different styles, the most important elementary concept has to do with the type of steps. A Tsamiko dance consists of multiple dance cycles, each one consisting of ten distinct steps. Each step is characterized and distinguished from the other steps by its four attributes (i.e. left or right direction, left or right foot, single or double step and foot is in high or low position) and its place in the sequence (i.e. 1st, 2nd,...,10th). The attributes of some steps depend on the particular style of the dance being performed, while other step attributes remain the same for all styles. More specifically, among the step attributes that do not depend on the dance style, we identify the movement direction (i.e., left or right), as well as the body place (i.e., whether it is the right or the left foot). Particularly, the first six steps have a "right" direction and the rest have a "left" direction. Also, 1st, 3rd, 5th, 8th, 10th are performed with the right foot and the rest with the left foot. Another important attribute that now depends on whether the style is "male" or "female", has to do with the foot lifting movement, which essentially differentiates between a step that is performed with the foot in high position, or in a position close to the floor. More specifically, the foot is high at the 6th and 10th step in a "male" dance while it is always low for "female" style. Also, in a dance of "double" style, the 2nd, 4th and 8th standard step of the dance cycle are characterized by the "double" attribute. On the other hand , all steps have the "single" attribute when the style is "single". Finally, the ontology of Figure 1 reveals also the importance of sequence among the undertaken steps that has to be performed in a rather strict order. Thus, it is evident that the detection of each step along with its attributes is crucial for our analysis framework.

6.2 Sensor signal analysis for elementary concept detection

In order to capture the performance of the dancers, we have used markerless motion capture based on depth sensing technology. Microsoft Kinect sensors were employed, which are low-cost real-time depth sensing cameras that can track the volume of a performer and produce skeletal data. Microsoft Kinect SDK [16] has been used as a solution for skeletal tracking and acquisition. It provides the ability to track the 3D positions of 20 predefined skeletal joints of a human body at 30Hz rate. In order to solve occlusion and self-occlusion tracking problems inherent in this type of motion capture and to increase the total area of coverage, several Kinect sensors were placed in an array in front of the dancer (Figure 2 Left). The captured data were combined following a fusion strategy described in [17], leading to an increased robustness of skeletal tracking.

The elementary domain concepts, which are the steps and the way they are executed, were extracted from the analysis of the joint position signals. The analysis consists of two parts: segmentation and feature extraction. Segmentation is performed on two levels of granularity. Initially, the dance periods are detected. Tsamiko dance has a repeating pattern, with the dancer moving on a semi-circle performing several steps to the right direction followed by several steps to the left. This consists of a single period, which is easily detected by analyzing the position of the waist of the dancer. Peak detection of a sub-sampled (to remove noise) waist displacement along the horizontal axis reveals the time instants when the dancer is at the end of the left/right movement (Figure 2 Right). Subsequently, each period is further segmented into steps which we consider an elementary domain concept. The detection of steps is based on the movement of ankles along the horizontal axis relative to the movement of the root of the body. Once again, local maxima detection is employed for the detection of time instants when the footstep is performed (when the foot touches the ground, the relative horizontal displacement produces local maxima).

After the segmentation, each segment is analyzed to extract the features of each step. The features extracted from each step are: the foot that is moving (left foot or right foot), the direction of movement (left or right), raised foot and double step. Those features are extracted from a rule based analysis of ankle and knee joint position signals of the dancers' legs. The double step is a sequence of two small steps executed sequentially, which we consider as a single step during the segmentation period, since the intermediate steps are small and executed very quickly. The result of this analysis is a sequence of steps together with properties assigned to each step which are used subsequently to infer the dance style.

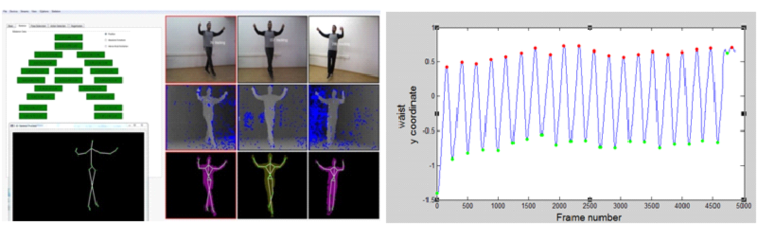

Fig. 2. (Left) Tsamiko performance captured by three depth cameras. Skeletal tracking from each camera can be seen as well as the final fused skeleton tracking result. (Right) Displacement of the waist of the dancer along the horizontal axis. Red and green dots represent the peaks and valleys detected, segmenting the dance into periods.

6.3 MTheory and MFrags

Based on the ontology described in Section 6.1, we have developed the MEBN of Figure 3. In this figure, a MEBN is presented consisting of two MFrags. The TsamikoStepMFrag contains information about the step sequence for different dance styles, along with the style distinguishing characteristics of the steps. The TsamikoStyleMFrag contains the style related MEBN nodes, “genderstyle” and “stepstyle” that can take the values male/female and single/double, respectively. Each MFrag contains input nodes (colored in grey), resident nodes (colored in yellow) and logical nodes (colored in green). The input nodes of the TsamikoStepMFrag are resident nodes in the TsamikoStyleMFrag. The only exception is the node “step”, which is used to model a recursive process as described below.

The “step” input node in TsamikoStepMFrag (colored in grey) models the step sequence that comprise a dance cycle. It is very important to understand the concept of recursion in MEBNs, which is manifested in this case by making the input (grey) node "step" the parent of the resident (yellow) node "step" (both having the same name but different logical variables ($t1$ and $t2$) as arguments). These variables are logical variables that are used to model the aspect of time sequence in the detected steps. For example, if $t2$ is instantiated as $timeStep_i$, then, based on the constraints dictated by the logical (green) nodes, $t1$ is instantiated as $timeStep_{i-1}$. In this way, $t1$ is always the previous time step of $t2$ enforcing the desired recursive process. In our example of Figure 3, the resident (yellow) node "step" has a range set of ten values, TsamikoStep1, ..., TsamikoStep10, each value denoting one of the ten distinct steps in a Tsamiko dance cycle. Thus, with the recursive definition of the node "step" (i.e. both as an input and a resident node) we enable the modeling of a dance step sequence execution. Note that the number of cycles and the starting and ending step are arbitrary.

Besides "step", there are four more resident nodes in the TsamikoStepMFrag as depicted in Figure 3: "hasDirection", "foot", "isFootHigh" and "isDoubleStep", which are essentially the features that declare the execution method of each step. The first node can take either the "leftDirection" or the "rightDirection" value, while the second node can take the "leftFoot" or "rightFoot" values. Both are not directly dependent to the dance style, as shown by the lack of direct arrows between these nodes and the nodes "genderstyle" and "stepstyle". Instead, the impact of "hasDirection" and "foot" to the recognition of the dance style goes through the "step" node that models the execution pattern of the dance. On the other hand, the nodes "isFootHigh" and "isDoubleStep" take boolean values and directly depend on the nodes "genderstyle" and "stepstyle" that determine the dance style. These dependencies are better described in the following paragraph that explains the TsamikoStyleMFrag.

According to the ontology presented in Section 6.1, the style of Tsamiko dance can be characterized as *male* or *female* and as double or single. We have decided to recognize the undertaken style on a per step-basis, meaning that steps of the same sequence can be attributed to different styles. Nevertheless, there is strong correlation between the style detected in $step_i$ and the probability that $step_{i+1}$ will follow the same style. Thus, as in the case of steps, there is an inherent requirement for modeling a recursive relation between the variables determining the style characteristics. To this end, the TsamikoStyleMFrag consists of two variables "genderstyle" and "stylestep" that exist both as resident and input nodes in the same MFrag, modeling the recursive relation between the styles detected for each step. Similar to TsamikoStepMFrag, the TsamikoStyleMFrag consists also of the exact same logical variables $t1$ and $t2$ that are instantiated to time steps and are used to enforce the desired recursive process. Finally, we should note that "genderstyle" and "stepstyle" are also used as input nodes in the TsamikoStepMFrag and act as direct parents of "isFootHigh" and "isDoubleStep".

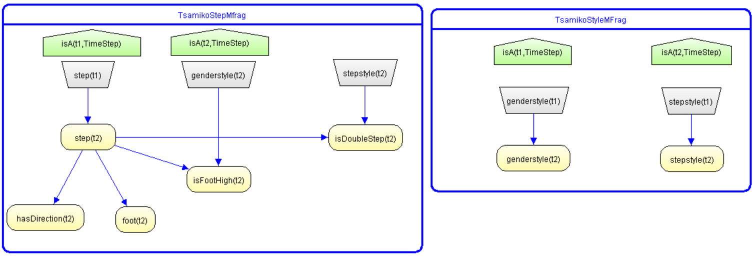

Fig. 3. MEBN developed to facilitate the analysis of greek tsamiko dance.

Table 1. Characteristics of the different performances used for the tsamiko dance.

performance	length	attributes	performance	length	attributes
A1	48	single/female	A9	201	double/male
A2	153	single/female	A10	190	double/female
A3	194	single/male	A11	198	double/female
A4	196	single/female	A12	189	double/female
A5	189	single/female	A13	202	double/male
A6	200	single/male	E1	190	single/male
A7	202	single/male	E2	100	double/female
A8	195	single/female	E3	191	double/male
A9	201	double/male	-	-	-

7 Preliminary Experimental Results

The goal of our experimental study is to verify the appropriateness of MEBNs in recognizing the different dance styles based on the undertaken steps. Actually, our interest is not in just classifying a step sequence to one of the existing dance styles, which would constitute a trivial problem. Instead, our goal is to classify each step to one of the existing dance styles and at the same time assess the proficiency level of the performer. In the first case we expect that the classification accuracy of the MEBN-based framework will outperform a baseline approach that relies on BNs but does not make any use of the situation specific capability of MEBNs. In the second case, we expect that our MEBN-based analysis framework will rank high the step sequences that have been performed flawlessly and rank low the step sequences that contain one or more execution errors.

7.1 Dataset and evaluation metrics

In our experiments we have used 16 recorded performances of Tsamiko dance, with the sequence length ranging from 50 to 200 steps. Out of the 16 recorded performances 3 were executed by professional dancers (E1-E3) while the rest was obtained from apprentice level dancers (A1-A13). All performances were executed with the same musical piece and every performance was annotated with its dance style, as depicted in Table 1.

In order to assess the classification accuracy we apply a threshold on the confidence score extracted for each step. More specifically, when we analyze a step sequence that is annotated as "single" we consider as correctly classified all steps that have caused the posterior probability (i.e. confidence degree) of the "step-style" node to overcome the 0.5 threshold. Similar is the case when analyzing a step sequence annotated with the other style types. Then, we divide this number with the total number of steps so as to calculate the classification accuracy for the entire performance. On the other hand, in order to assess the proficiency level of each performance we use the average of the confidence degrees of all steps in this performance. More specifically, the result of the analysis process for each performnance is a two-dimensional table with its columns corresponding to the different styles, its rows corresponding to the individual steps and its values being the posterior probability of the SSBN random variables modeling the dance style information. Thus, by performing column-wise averaging in this table we obtain four scores (i.e. corresponding to the four dance styles) that are suitable for assessing the proficiency level of the undertaken step sequence (i.e. considering that the closer you get to a 100% score the closer you get to a flawless performance).

7.2 Step classification accuracy

In order to verify the benefit of being able to adapt to the situation at hand, we compare our MEBN-based model with a baseline approach that lacks this capability. More specifically, given that one of the auxiliary features offered by the signal analysis module is the total number of steps composing the step sequence, the baseline approach was designed to totally neglect this situation specific information. It is essentially implemented as a straightforward BN with a fixed number of nodes representing steps (we have used 10 steps which is the standard cycle in a tsamiko dance) and without any provision for the recursive relation between steps and dance cycles. Figure 4 demonstrates the classification accuracy of both frameworks. We can see that the MEBN-based framework outperforms the baseline approach in 28 out of the 32 cases. Given that both frameworks rely on probabilistic inference and both frameworks have been designed based on the same domain knowledge, it is reasonable to attribute the observed improvement in the flexibility of the MEBN-based framework to adapt in the number of steps composing each performance.

7.3 Proficiency level assessment

Figure 5 shows the proficiency level results for the 16 performances, grouped based on the dance style and distinguishing between apprentices and experts. Moreover, apart from the average score the standard deviation is also depicted. The obtained results verify our expectation that in the majority of the examined cases our MEBN-based analysis framework is able to distinguish between an apprentice and an expert. This is evident in the case of "female" style where the performance level of the expert is higher than all apprentices. Similar conclusions

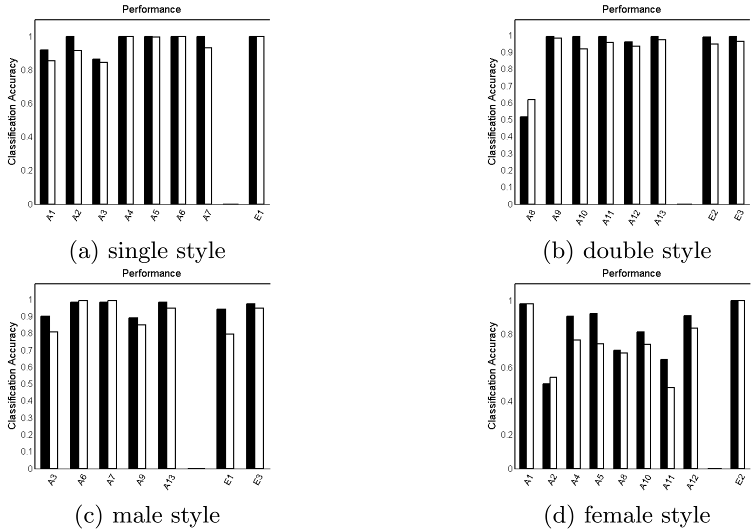

Fig. 4. The performance results are grouped based on the style. The bar diagrams colored in black correspond to the MEBN-based framework, while the bar diagrams colored in white correspond to the baseline.

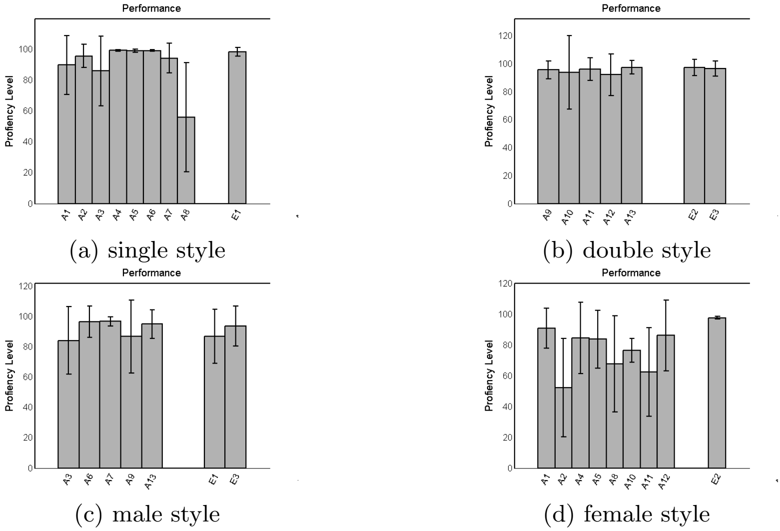

Fig. 5. Proficiency level results for the 16 performances of our dataset, grouped based on the dance style and distinguishing between apprentices and experts.

can be also extracted for the cases of “single” and “double” where despite the fact that the proficiency level of the experts does not supersede all apprentices their superiority is evident in terms of average numbers. Finally, in the case of “male” style we notice that the proficiency level of the experts is lower by approximately 1.5% than the average score of all apprentices. This outcome can be attributed to the low performance of our MEBN-based analysis framework in effectively modeling one of the key-features characterizing the “male” style, which is the detection of “isFootHigh”. Since the detection of this feature is rather challenging for the signal analysis module, it seems that in this case our MEBN-based model has failed to prevent the propagation of the first stage analysis error to the final outcome. In our future work we plan to examine the score of each step in correlation with the performance of the signal analysis module so as to gain more insights.

8 Conclusions

In this paper we have shown how the theory of MEBNs can be used to combine probabilistic analysis with first-order logic. The proposed framework was employed for the semantic analysis of Tsamiko traditional dances. The purpose of semantic analysis was to recognize the specific style of the tsamiko dance based on the special characteristics of the dance steps. The latter were extracted by a motion analysis module relying on the body movements. Experiments demonstrated that the classification efficiency of the proposed model is significantly better than the standard Bayesian network case. Also, the model was evaluated in terms of the ability to discriminate between expert and apprentice dancers, giving encouraging results. In the future, we plan to augment the MEBN model with information obtained from other than visual based modalities, such as sound. Precisely, we expect that by exploiting the information from the musical piece accompanying the dance performances, we can improve the accuracy of semantic analysis. However, the task of combining music and body movement information is challenging, since it requires the identification of the musical features that provide useful information (i.e., that have a semantic meaning for the dance) and the detection of their dependency with the dance steps.

Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7-ICT-2011-9) under grant agreement no FP7-ICT-600676 “i-Treasures: Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures”.

References

1. Biederman, I.: Recognition-by-components: A theory of human image understanding. Psychological Review **94** (1987) 115–147
2. Li, F.F., Fergus, R., Perona, P.: One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4) (2006) 594–611
3. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. IEEE Trans. Pattern Anal. Mach. Intell. **30**(6) (2008) 985–1002
4. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 1958–1970
5. Laskey, K.B.: Mebn: A language for first-order bayesian knowledge bases. Artificial Intelligence **172** (2008) (2–3), 140178
6. Mallik, A., Chaudhuri, S., Ghosh, H.: Nrityakosha: Preserving the intangible heritage of indian classical dance. ACM Journal on Computing and Cultural Heritage **4** (2011) (3), 11
7. Koller, D., Pfeffer, A.: Object-oriented bayesian networks. In: Uncertainty in Artificial Intelligence: Proceedings of the Thirteenth Conference, San Francisco, CA: Morgan Kaufmann. (1997)
8. Pfeffer, A.: Probabilistic Reasoning for Complex Systems. Stanford, CA, Stanford University (2000)
9. Mallik, A., Pasumarthi, A.P., Chaudhury, S.: Multimedia ontology learning for automatic annotation and video browsing. In: 1st ACM International Conference on Multimedia Information Retrieval (MIR'08), New York,. (2008)
10. Davis, E.: Representations of Commonsense Knowledge. San Mateo, California: Morgan Kaufmann (1990)
11. Cardoso, J.: The semantic web vision: Where are we? IEEE Intelligent Systems **22**(5) (2007) 84–88
12. Horrocks, I.: Description logics in ontology applications. In: Automated Reasoning with Analytic Tableaux and Related Methods. (2005) 2–13
13. Ding, Z., Peng, Y., Pan, R.: A bayesian approach to uncertainty modeling in owl ontology. In: International Conference on Advances in Intelligent Systems - Theory and Applications. (November 2004)
14. Whittaker, J.: Graphical models in applied multivariate statistics. Wiley Publishing (2009)
15. Suárez-Figueroa, M.C., Gómez-Pérez, A., Villazón-Terrazas, B.: How to write and use the ontology requirements specification document. In: Proceedings of the Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009 on On the Move to Meaningful Internet Systems: Part II. OTM '09, Berlin, Heidelberg, Springer-Verlag (2009) 966–982
16. Microsoft: Kinect for windows — voice, movement and gesture recognition technology. `http://www.microsoft.com/en-us/kinectforwindows/` (2013)
17. Kitsikidis, A., Dimitropoulos, K., Douka, S., Grammalidis, N.: Dance analysis using multiple kinect sensors. In: VISAPP2014. (2014)