Category of Markov Systems

Vladimír Marek*
marek@dcse.fee.vutbr.cz

Abstract: Describing mathematical models in terms of category theory can help us understand them better, find new interesting properties and relations to other mathematical models. This paper studies Markov systems as models of stochasticity, shows one of their possible category, and presents its categorical properties.

Key Words: Category, morphism, Markov system, Markov chain, probability.

1 Introduction

Markov systems along with Markov chains are fundamental models of transition systems with probabilistic behaviour. A common property of these models is that they can describe a transition system (having finite or at most countable set of states) and probabilities of transitions between the states. These two types of models differ in how transition probability between two given states can be changed:

a) *homogeneous Markov chains* – the transition probability between two given states is constant, time-invariant
b) *inhomogeneous Markov chains* – the transition probability between two given states is variable, time-dependent
c) *Markov systems* – the transition probability between two given states is variable, dependent on an input symbol

Markov systems can be viewed from more viewpoins – they can be regarded as stochastic automata reacting to its input, or unreliable deterministic automata (unreliability is expressed by decreasing some transition probabilities from 1 to lower values, along with adding some new "error" transitions with nonzero probability), or as generators of Markov chains (given an fixed input word to a Markov system, we obtain an, in general, inhomogeneous Markov chain).

Before we state a formal definition of Markov systems, we have to explain some terms used throughout the text (precise definitions in [1]).

- *Stochastic vector* – row vector $v = [v_1, v_2, ..., v_n]$, entries of which are real numbers and satisfy two conditions:
  1) $0 \le v_i \le 1$,
  2) $\Sigma v_i = 1$

  A stochastic vector describe probability distribution over states of a system (v_i is interpretable as a probability of the system being in state i).
- *Stochastic matrix* – square matrix, rows of which are stochastic vectors. A stochastic matrix can describe conditional probabilities of each *transition*, or, probability of the

* Department of Computer Science and Engineering, Božetěchova 2, CZ-612 66 Brno

system going to state *j* given it had been in state *i*. Product of stochastic matrices (if defined) is a stochastic matrix (proof in [1]).

- *General stochastic matrix* – square or non-square matrix, rows of which are stochastic vectors
- *Degenerate stochastic matrix/vector* – stochastic matrix/vector entries of which are 0 or 1 only
- *Characteristic matrix of a mapping* – given mapping $f: S \to S'$ where S, $S' \neq \emptyset$, $S = \{s_1, s_2, ..., s_n\}$, $S' = \{s_1', s_2', s_n'\}$, its characteristic matrix $M = [m_{ij}]$ is defined as a matrix of dimension $n \times n'$ where $m_{ij} = \begin{cases} 1 \text{ if } f(s_i) = s_j' \\ 0 \text{ if } f(s_i) \neq s_j' \end{cases}$. It is obvious that if a matrix is a characteristic matrix of a mapping, then it is degenerate general stochastic.

Definition 1.1 – Markov system

A *Markov system* is a triple $(S, \Sigma, \{P(\sigma_i)\})$, where S is a nonempty, finite or at most countable set of states, Σ is a nonempty, finite alphabet, and $\{P(\sigma_i)\}$ is a set of Markov matrices representing the transitions between states such that the matrix $P(\sigma_i)$ is associated to the symbol $\sigma_i \in \Sigma$.

If x is a word in Σ^* (which is the set of all words over Σ including the empty word denoted by Λ) such that $x = \sigma_1\sigma_2 \ ... \ \sigma_k$ then $P(x) = P(\sigma_1)P(\sigma_2) \ ... \ P(\sigma_k)$; $P(x) = [p_{ij}(x)]$ and $p_{ij}(x)$ is a transition probability from state i to state j associated with the word x. A matrix associated with the empty word is the identity matrix: $P(\Lambda) = \boldsymbol{1}$.

A Markov system, the alphabet of which consists of one element ($|\Sigma| = 1$) is the same thing as an homogeneous Markov chain. The chief difference between Markov chains and Markov systems is that by choosing different sequences from Σ^* we can generate many chains within a system.

It is clear that all $P(\sigma_i)$ and $P(x)$ for any word x are stochastic matrices (proof in [1]).

2 Category of Markov Systems – MSys

Firstly, we specify homomorphisms between Markov systems and then we define the desired category.

Definition 2.1 – Homomorphisms between Markov systems

Let $A = (S, \Sigma, \{P(\sigma_i)\})$ and $A' = (S', \Sigma', \{P'(\sigma_j')\})$ be two Markov systems. A homomorphism $h: A \to A'$ is a pair (m, l) of mappings:

$m: S \to S'$

$l: \Sigma \to \Sigma'$

which satisfies the following intertwining condition:

$$\forall \sigma_i \in \Sigma: \qquad M P'(l(\sigma_i)) = P(\sigma_i) M$$

where M is a characteristic matrix of m for some ordering of S and S'; it is a $|S| \times |S'|$ general degenerate stochastic matrix.

Since each mapping can be described in terms of its characteristic matrix, an alternative specification of the homomorphism can be pair (M, l).

Note: this condition has been found to preserve probability between two systems. The probability-preserving law is not studied here since it exceeds the range of this paper.

In case of the alphabet being fixed, l is the identity function over Σ; the condition is then simplified,

$$\forall \sigma_i \in \Sigma: \qquad M P'(\sigma_i) = P(\sigma_i) M.$$

Definition 2.2 – Category of Markov systems – MSys

The category of Markov systems has the following objects and mappings.

- *Objects* are all Markov systems.
- *Mappings* are all homomorphisms between Markov systems.

The *composition* of mappings is defined as element-by-element composistion of functions, i.e. for every composable mappings $f = (m_f, l_f)$: $A \to B$ and $g = (m_g, l_g)$: $B \to C$ the composition mapping $g \cdot f$ is defined:

$$g \cdot f = (m_g \cdot m_f, l_g \cdot l_f)$$

The associative law holds in the category of sets, so the law come true in this category either. Unlike in the category of sets, in this category there is a restriction that both domains and codomains of m's and l's can not be empty sets, which follows the definition of Markov systems (S and Σ must not be empty).

If we use the alternative definition of homomorphisms $f = (M_f, l_f)$: $A \to B$ and $g = (M_g, l_g)$: $B \to C$, the composition is defined:

$$g \cdot f = (M_f * M_g, l_g \cdot l_f)$$

Here the element-by-element composition of state mapping m is substituted by matrix multiplication. The associativity law holds since matrix multiplication is associative.

The composition of the homomorphisms is also a homomorphism since the homomorphism condition for the composition is satisfied.

Proof.

Let $A = (S, \Sigma, \{P(\sigma_i)\})$ and $A' = (S', \Sigma', \{P'(\sigma_j')\})$ and $A'' = (S'', \Sigma'', \{P''(\sigma_j'')\})$ be three Markov systems, and $f = (M_f, l_f)$ and $g = (M_g, l_g)$ two morphisms such that:

$$A' = f(A)$$

$$A'' = g(A')$$

Obviously, this satifies two conditions:

(1) $\forall \sigma_i \in \Sigma: \qquad M_f P'(l_f(\sigma_i)) = P(\sigma_i) M_f$

(2) $\forall \sigma_j \in \Sigma': \qquad M_g P''(l_g(\sigma_j')) = P'(\sigma_j') M_g$

The composition $g \cdot f$ is then required to satisfy

(3) $\forall \sigma_i \in \Sigma: \qquad (M_f * M_g) P''((l_g \cdot l_f)(\sigma_i)) = P(\sigma_i) (M_f * M_g)$

The left-hand side of this equation can be adjusted in the following way

$$M_f M_g P''(l_g(l_f(\sigma_i))) =$$

$= M_f P'(l_f(\sigma_i)) M_g$ because of condition (2), and

$= P(\sigma_i) M_f M_g$ because of condition (1), so condition (3) holds true.

∎

The *identity mapping* of this category is morphism $\mathbf{id} = (\mathbf{id}_S, \mathbf{id}_\Sigma)$ such that:

$\mathbf{id}_S: S \to S: \mathbf{id}_S(x) = x$

$\mathbf{id}_\Sigma: \Sigma \to \Sigma: \mathbf{id}_\Sigma(x) = x$

The identity mapping signifies the matrix M being unit; $M = \boldsymbol{1}$ (a characteristic matrix of the identity mapping is the identity matrix). Now we check the equivalency of $A = ((S, \Sigma, \{P(\sigma_i)\})$ and $A' = ((S', \Sigma', \{P'(\sigma_i')\},)$, such that $A' = \mathbf{id}(A)$.

$\forall \sigma_i \in \Sigma: \qquad M P'(l(\sigma_i)) = P(\sigma_i) M,$

and since $l = \mathbf{id}_\Sigma$ and $M = \boldsymbol{1}$,

$\boldsymbol{1} P'(\mathbf{id}_\Sigma(\sigma_i)) = P(\sigma_i) \boldsymbol{1}$

$P'(\sigma_i) = P(\sigma_i).$

Besides $S = S'$ and $\Sigma = \Sigma'$ due to $\mathbf{id}_S$ and $\mathbf{id}_\Sigma$ being identity mappings over S, or Σ, respectively. Hence $A = A'$, or $A = \mathbf{id}(A)$ for any object A of this category.

Finally we have checked all the properties that every category must satisfy, so the Markov systems and their homomorphisms form the category.

3 Categorical Properties of MSys

3.1 Epimorphisms and Monomorphisms

If $f = (m_f, l_f)$, $g = (m_g, l_g)$, and $h = (m_h, l_h)$ then the problems of finding the morphisms f satisfying $g \cdot f = h \cdot f \Rightarrow g = h$ (epimorphisms), or satisfying $f \cdot g = f \cdot h \Rightarrow g = h$ (monomorphisms) can be decomposed to two subproblems which must hold true at the same time; therefore we use logical *and* between the two conditions:

$$(m_g \cdot m_f = m_h \cdot m_f \Rightarrow m_g = m_h) \wedge (l_g \cdot l_f = l_h \cdot l_f \Rightarrow l_g = l_h)$$

in case of f being an *epimorphism*, and

$$(m_f \cdot m_g = m_f \cdot m_h \Rightarrow m_g = m_h) \wedge (l_f \cdot l_g = l_f \cdot l_h \Rightarrow l_g = l_h)$$

if we investigate whether or not f is a *monomorphism*.

This yields that Markov system homomorphism $f = (m_f, l_f)$ is an *epimorphism* if and only if either its component is surjective. Similarly, the homomorphism f is a *monomorphism* if and only if either its component is injective.

If we use the 'matrix-like' definition of homomorphisms, $f = (M_f, l_f)$ is an *epimorphism* if and only if l_f is surjective and M_f does not contain any non-zero column. The homomorphism f is a *monomorphism* if and only if l_f is injective and each column of M_f contains at most one entry 1. The proof of the required properties of M_f is not given since it is obvious if we consider that M_f reflects the mapping m_f; it is its characteristic matrix.

3.2 Products

Let $A_1 = (S_1, \Sigma_1, \{P_1(\sigma_{i1})\})$ and $A_2 = (S_2, \Sigma_2, \{P_2(\sigma_{j2})\})$ be two Markov systems. Then the product $A_1 \times A_2$ is a system

$A_1 \times A_2 = (S_1 \times S_2, \Sigma_1 \times \Sigma_2, \{P_{12}(\sigma_{ij12})\})$ where

$P_{12}(\sigma_{ij12}) = P_1(\sigma_{i1}) \otimes P_2(\sigma_{j2})$; in other words $\{P_{12}(\sigma_{ij12})\}$ is the outer Kronecker product $\{P_{12}(\sigma_{ij12})\} = \{P_1(\sigma_{i1})\} \times \{P_2(\sigma_{j2})\}$

The projections $\pi_1 = (m_{\pi 1}, l_{\pi 1})$: $A_1 \times A_2 \to A_1$ and $\pi_2 = (m_{\pi 2}, l_{\pi 2})$: $A_1 \times A_2 \to A_2$ are defined as follows:

$$m_{\pi 1} = \{(s_1, s_2) \mapsto s_1\} \qquad m_{\pi 2} = \{(s_1, s_2) \mapsto s_2\}$$
$$l_{\pi 1} = \{(\sigma_1, \sigma_2) \mapsto \sigma_1\} \qquad l_{\pi 2} = \{(\sigma_1, \sigma_2) \mapsto \sigma_2\}$$

The characteristic matrix $M_{\pi 1}$ has dimensions $|S_1||S_2| \times |S_1|$, and $M_{\pi 2}$ has dimensions $|S_1||S_2| \times |S_2|$. Providing that the states in the product are ordered so that (s_i, s_j) comes before (s_k, s_l) iff $i < k$ or $i = k \wedge j < l$ which means lexicographically, the matrix $M_{\pi 1}$ consists of $|S_1|$ (numbered by q, $1 \le q \le |S_1|$) matrices $|S_2| \times |S_1|$, one over another, each of which has all 1's in q-th column; while $M_{\pi 2}$ consists of $|S_1|$ unit matrices of dimension $|S_2|$, one over another. Graphically,

$$M_{\pi 1} = \left[\begin{array}{ccc} 1 & & \\ \vdots & & 0 \\ 1 & & \\ \hline & 1 & \\ 0 & \vdots & 0 \\ & 1 & \\ \hline & \vdots & \\ \hline & & 1 \\ & 0 & \vdots \\ & & 1 \end{array}\right], \quad M_{\pi 2} = \left[\begin{array}{ccc} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ \hline 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ \hline & \vdots & \\ \hline 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{array}\right].$$

If we choose any $\sigma_{i1} \in \Sigma_1$, $\sigma_{j2} \in \Sigma_2$ and denote $P_1 = P_1(\sigma_{i1})$, $P_2 = P_2(\sigma_{j2})$ these equations hold:

(1) $M_{\pi 1} * P_1 = (P_1 \otimes P_2) * M_{\pi 1}$, and

(2) $M_{\pi 2} * P_2 = (P_1 \otimes P_2) * M_{\pi 2}$, and

We show the proof only for (1); the proof for (2) is similar.

Proof. (1)

Let us denote the matrix $(P_1 \otimes P_2) * M_{\pi 1}$ as V; $V = [v_{ij}]$,

$P_1 = [{}^1p_{ij}]$, $P_2 = [{}^2p_{ij}]$, $P_1 \otimes P_2 = [{}^{12}p_{ij}]$, $M_{\pi 1} = [{}^1m_{ij}]$

$dim(P_1) = n_1$, $dim(P_2) = n_2$, $dim(P_1 \otimes P_2) = n_1 n_2$, $M_{\pi 1}$ has dimension $n_1 n_2 \times n_1$ as well as V.

To simplify the expression, we will use two indexing systems: r, s, and i, j, k, l. We can recalculate them with using these formulas:

$$r = dim(P_2) * (i - 1) + k$$
$$s = dim(P_2) * (j - 1) + l$$

Each element of V can be expressed

$$v_{rq} = \sum_{s=1}^{n_1 n_2} {}^{12}p_{rs}\, {}^1m_{sq} = \sum_{s=1}^{n_1 n_2} {}^1p_{ij}\, {}^2p_{kl}\, {}^1m_{sq} = \sum_{l=1}^{n_2} \sum_{j=1}^{n_1} {}^1p_{ij}\, {}^2p_{kl}\, {}^1m_{n_2(j-1)+l,q}\,.$$

To continue with further derivations we must realise that

$$m_{n_2(j-1)+1,q} = \begin{cases} 1 \text{ if } j = q \\ 0 \text{ if } j \ne q \end{cases}$$

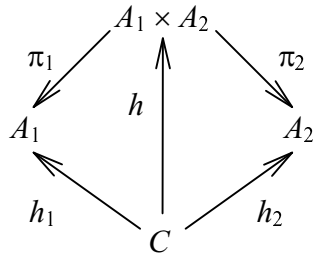
Therefore

$$v_{rq} = v_{n_2(i-1)+k,q} = \sum_{l=1}^{n_2} {}^1p_{iq}\,{}^2p_{kl} = {}^1p_{iq} \sum_{l=1}^{n_2} {}^2p_{kl} = {}^1p_{iq}$$

because P_2 is a stochastic matrix.

∎

But to say that this is a product in category MSys, we have to prove that the following product definition diagram commutes and for given A_1, A_2, C and h_1, h_2 there exists a unique h.



Proof.

The unique mapping h: $C \to A_1 \times A_2$ is defined

$h = (h_1(x), h_2(x))$.

By the definition of projections, h is unique since h_1 is forced to be the first component and h_2 the second. Only this case can make the diagram commute.

$G * M_{\pi 1} = G_1$ is true since $(G_1 \otimes G_2) * M_{\pi 1} = G_1$ is true

$G * M_{\pi 2} = G_2$ is true since $(G_1 \otimes G_2) * M_{\pi 2} = G_2$ is true

where $h = (G, l)$, $h_1 = (G_1, l_1)$ and $h_2 = (G_2, l_2)$.

∎

4 Conclusion

Although we have found the category of Markov systems and specified its monomorphisms, epimorphisms and products precisely, we have not yet been successful in finding (or proving that they do not exist) sums and other categorical properties. Therefore this work is still open.

This work has been done within the research intention No. CEZ: J22/98: 262200012 – "Research in Information and Control Systems" and supported by the Grant Agency of Czech Republic grant No. 102/98/0552 "Research and Application of Heterogeneous Models".

References

1. Marek, V.: *Categorical Approach to Markov Systems*. MSc Thesis (unpublished), Brno University of Technology, 1999.
2. Arbib, M. A., Manes, E. G.: *Arrows, structures, and functors: The categorical imperative*, Academic Press, New York, San Francisco, London, 1975.
3. Lawvere, F. W., Schanuel, S. H.: *Conceptual mathematics: A first introduction to categories*, Cambridge University Press, 1997.
4. Paz, A.: *Introduction to probabilistic automata*, Academic Press, New York and London, 1971.