2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Associative Processes between Behavioral Symbols and a Large Scale Language Model

Wataru Takano and Yoshihiko Nakamura

Abstract— **This paper describes an novel approach towards linguistic processing for robots through integration of a motion language model and a natural language model. The motion language model works for association of words from motion symbols. The natural language model is one used for a morphological analysis, which has been developed in natural language community. The natural language model is optimized using a enormous amount of words. So this model is scalable architecture. The motion language model and the natural language model can be integrated since both models are represented graphically. The integration of the motion language model and the natural language model allows robots not only to interpret motion patterns as sentences but also to generate motions from sentences. This paper demonstrates the validity of our proposed framework even in the case that large-scale word corpus is needed processing through experiments of interpreting motion patterns as sentences and generating motion patterns from sentences.**

I. Introduction

Language is an indispensable symbolic system to representation of knowledge, communication and reasoning. Language acquisition of a robot is required for its intelligence, but it is extremely tough problem.

In robotics, some approaches for symbolization of robot's motion patterns have been proposed, such as MOdule Selection And Identification for Control (MOSAIC) [1], neural network approach [2][3], and Hidden Markov Model approach (HMM) [4][5]. Additionally a model of nonverbal communication between a humanoid robot and its partner based on the symbolization of motion patterns is presented [6]. Although the communication model consists of two hierarchies of symbolic representations of motion patterns and behavioral interaction patterns, the communication is not based on linguistic processing. Sugita et al. introduced a connectionist model, which consists two Recurrent Neural Networks with Parametric Bias (RNNPBs) [7]. One RNNPB symbolizes motion patterns and another one represents a simple finite language. The two RNNPBs are combined through Parameteric Bias layer. A robot can generate a motion from a sentence. Ogata et al. additionally proposed a method to generate a sentence from a motion by using the same framework as Sugita's model [8]. These framework has two drawbacks of scalability and ambiguity. A connectionist model is not suitable to learn a large-scale language corpus. A motion can be expressed by various kinds of sentences.

In the field of video analysis, a number of approaches to extract image features using HMMs have been proposed [9]. In the natural language processing, a probabilistic content model has been proposed to represent topics and topic shifts [10]. They also used HMMs, where a state correspond to a topic. Shibata et al. has developed automatic topic identification based on HMMs with both image and language information [11]. This framework stochastically maps visual and audio features to topic words. Generation of the extracted features from the topic words is not realized.

We also have proposed stochastic approach to linguistic processing based on symbolization of motion for humanoid robots [12][13], which consists of motion symbol module and natural language module. Our stochastic approach can not only generate multiple motions from a sentence but also generate various kinds of sentences from a motion. However the natural language module learns only a small-scale language corpus. A large-scale language corpus has to be learned by the robots so that the robots make inferences in various ways based on language knowledge, which underlies in the corpus.

In a community of natural language processing, various kinds of natural language analysis have been developed. Especially, stochastic approaches are advantageous to processing with enormous quantity of linguistic data [14][15].

This paper describes a novel approach to linguistic processing based on two functions : semantic function between motion symbols and words, and grammatical function of sentences. The grammatical function is learned by a model of morphological analysis, which can deal with a large amount of words and word classes [14]. The integration of these two functions makes it possible for a robot not only to interpret motion patterns as sentences but also to generate motion patterns corresponding to sentences. This paper also verifies the validity of our proposed framework on experiments.

II. Motion Language Model and Natural Language Model

A. Motion Language Model

Although textual or phonetical modality included in language can be measured, some of the linguistic modalities which underlie the linguistic structure cannot be observed. In this paper, we propose a motion language model stochastically connecting symbols of motion patterns to morpheme words through latent variables, which are suitable to represent unobservable states. Fig.2 illustrates the motion language model.

The motion language model consists of three kinds of nodes : symbol of motion pattern, latent variable and morpheme word. The symbol of motion pattern is represented

W. Takano and Y. Nakamura are with Mechano-Informatics, University of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo, Japan, {takano, nakamura}@ynl.t.u-tokyo.ac.jp

978-1-4244-5040-4/10/$26.00 ©2010 IEEE

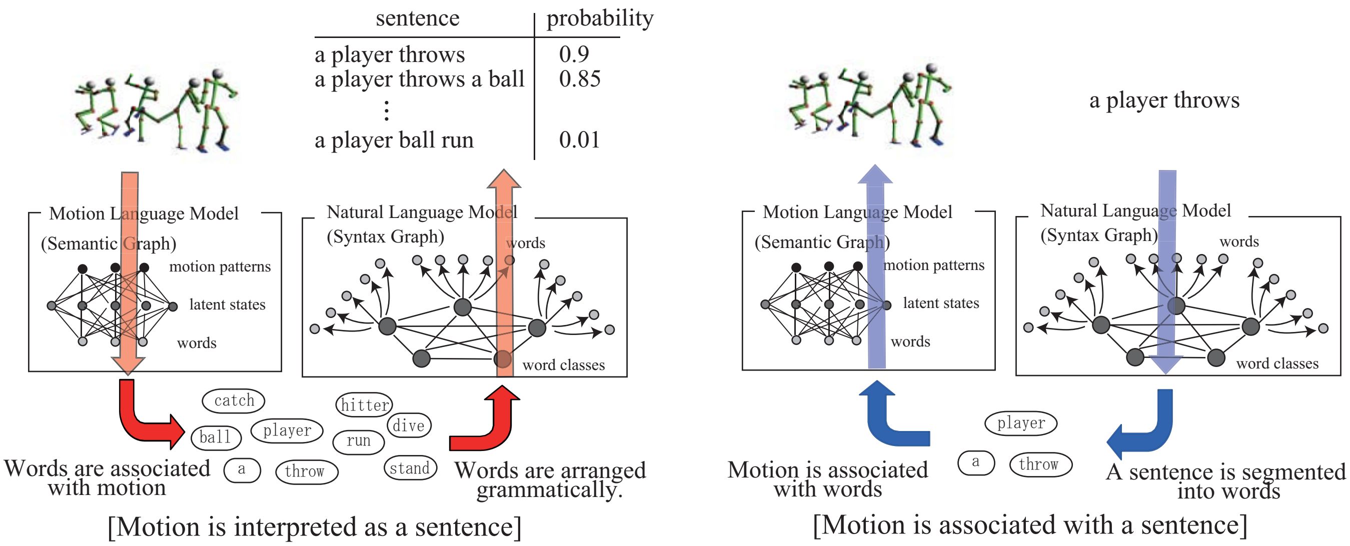


Fig. 1. Overview of integration of a motion language model with a natural language model. The motion language model represents relationship among motion symbols and morpheme words via latent variables as a graph structure, where nodes on 1st, 2nd and 3rd layer indicate the motion symbols, the morpheme words and the latent variables respectively. The natural language model represents the dynamics of language which means the order of words in sentences. The motion language model and the natural language model are equivalent to semantics and syntax. By integrating two functions, linguistic processing for robots can be realized.

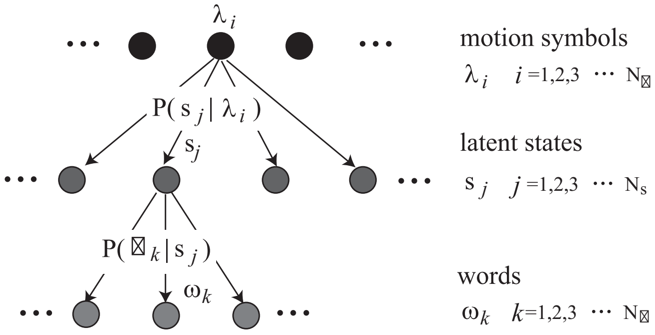


Fig. 2. The motion language model represent the stochastic association of Morpheme words with motion symbols via latent states. The motion language is defined by two kinds of parameters : probability that a morpheme word is generated by a latent variable and probability that a latent variable is generated by a motion symbol.

by an HMM which learns spatial and temporal behavioral pattern [16]. The connections between the symbol of motion pattern and the latent variable and between the latent variable and the morpheme word are expressed by associative probabilities : $P(s|\lambda)$ and $P(\omega|s)$. Note that λ, s and ω are the symbol of motion pattern, the latent variable and morpheme word respectively, and that $P(s|\lambda)$ and $P(\omega|s)$ represent the probability that the symbol of motion pattern λ generates the latent variable s and the probability that the latent variable s generates the morpheme word ω.

The stochastic parameters of the motion language model $P(s|\lambda)$ and $P(\omega|s)$ are optimized by EM (Expectation Maximization) algorithm, which alternately processes two steps : Expectation step (E-step) and Maximization step (M-step). Training pairs of symbol of motion pattern and a sentence (a sequence of morpheme words) are given. The training pair is described by $\left\{\lambda^k; \omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k k = 1, 2, 3, \cdots, N\right\}$. Note that N is the number of training pairs and that n_k is the number of the morpheme words composing the k-th sentence. Both of the symbol of motion pattern λ^k and the sentence $(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k)$ represent k-th motion pattern.

E-step calculates distributions of the latent variables based on model parameters estimated in previous M-step. The distributions of the latent variables are provided as follows.

$$P(s|\lambda^k, \omega_i^k) = \frac{P(\omega_i^k|s, \lambda^k, \theta)P(s|\lambda^k, \theta)}{\displaystyle\sum_{j=1}^{N_s} P(\omega_i^k|s_j, \lambda^k, \theta)P(s_j|\lambda^k, \theta)} \qquad (1)$$

where θ is a set of the previously estimated model parameters $P(s|\lambda)$ and $P(\omega|s)$.

M-step estimates the model parameters such that summation of expectation of log-likelihood that the symbol of motion pattern λ^k generates the sentence $(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k)$ is maximized. The summation of the expectation of the log-likelihood Φ is described by the following equation.

$$\Phi = \sum_{k=1}^{N} \log P(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k|\lambda^k) \qquad (2)$$

$$P(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k|\lambda^k) = \prod_{i=1}^{n_k} P(\omega_i^k|\lambda_k) \qquad (3)$$

$$P(\omega_i^k|\lambda^k) = \sum_{j=1}^{N_s} P(\omega_i^k|s_j)P(s_j|\lambda^k) \qquad (4)$$

where we uses conditional independence assumption expressed by Equation 3. The probability that a symbol of motion pattern generates a morpheme word can be rewritten as Equation 4. The estimates of the new model parameters

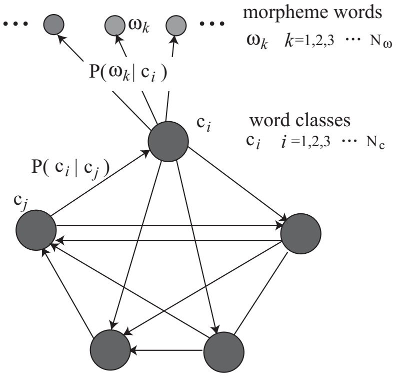


Fig. 3. Natural language model represents the dynamics of word classes by Hidden Markov Models. The node c_i corresponds to the word class. Transition from the node c_j to the node c_i is implemented with probability $P(c_i|c_j)$. The morpheme word ω_k is generated by the node c_i with conditional probability $P(\omega_k|c_i)$.

are as follows (see the appendix for how to derive the estimates).

$$P(s|\lambda) = \frac{\displaystyle\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\lambda,\lambda^k)P(s|\lambda^k,\omega_i^k)}{\displaystyle\sum_{j=1}^{N_s}\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\lambda,\lambda^k)P(s_j|\lambda^k,\omega_i^k)} \quad (5)$$

$$P(\omega|s) = \frac{\displaystyle\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\omega,\omega_i^k)P(s|\lambda^k,\omega_i^k)}{\displaystyle\sum_{j=1}^{N_\omega}\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\omega_j,\omega_i^k)P(s|\lambda^k,\omega_i^k)} \quad (6)$$

where $\delta(\lambda_i,\lambda_j)$ and $\delta(\omega_i,\omega_j)$ are Kronecker deltas. $\delta(\lambda_i,\lambda_j)$ and $\delta(\omega_i,\omega_j)$ become 1 if i is equal to j. Otherwise, they become 0 respectively. The numerators in Eqn.5 and Eqn.6 express expected number of times that hidden variable s is generated from motion symbol λ and expected number of times that hidden variable s is generated from word ω respectively. The denominators in Eqn.5 and Eqn.6 express the number of motion symbol λ in the training pairs and the expected number of times of hidden variable s in the training pairs.

By iteratively computing the distributions of the latent variables and the estimates of model parameters by using Equations 1, 5 and 6, we can derives the appropriate motion language model.

B. Natural Language Model

Various kinds of language models have been proposed in a community of natural language processing. Especially, stochastic models are advantageous such as CRF (Conditional Random Fields) [15] or HMM [14] since the linguistic model is required to deal with a lot of words. In this paper we use a morphological analysis model as a natural language model [14]. The morphological analysis model is represented by HMM as illustrates by Fig.3, where each node corresponds to a word class such as noun, verbs, adverb and so on. Note that words are classified in detail. For example, a verb word is classified depending to word inflexion. The node stochastically generates words that are classified to the node and the dynamics of the word classes are expressed by the stochastic transitions among the nodes. The natural language model is defined by a set of parameters: initial node distributions $\{\pi_i|i=1,,2,3,\cdots,N_c\}$ that initial morpheme words are classified as the word class c_i, the transition probabilities $\{P(c_i|c_j)|i,j=1,2,3,\cdots,N_c\}$ that the node c_i follows the node c_j, and the output probabilities $\{P(\omega_k|c_i)|i=1,2,3,\cdots,N_c,k=1,2,3,\cdots,N_\omega\}$ that the node c_i generates the morpheme word ω_k, where N_c indicates the number of nodes in the natural language model.

C. Combining Motion Language Model and Natural Language Model

Combining the motion language model and the natural language model enables robots to interpret motion patterns as sentences and to generate motion patterns from sentences. These two computational processes can be described by the stochastic searching problems.

1) Interpretation of Motion Patterns as Sentences: The interpretation of motion pattern as sentence can be made through recognition of a motion pattern as a symbol and form of a sentence corresponding to the symbol using the motion language model and the natural language model.

A motion pattern can be recognized as a symbol with the largest likelihood that the motion pattern is generated by a symbol. The motion recognition can be processed as follows.

$$\lambda^o = \arg\max_{\lambda_i:i=1,2,\cdot,N_\lambda} P(\boldsymbol{O}|\lambda_i) \quad (7)$$

where $P(\boldsymbol{O}|\lambda_i)$ is the likelihood that the motion pattern $\boldsymbol{O}$ is generated by the motion symbol λ_i, and the symbol λ^o is the result of motion recognition. Note that the motion pattern $\boldsymbol{O}$ is represented by temporal-spatial data such as a sequence of joint angles.

Composing a sentence corresponding to the symbol of motion pattern becomes searching a sequence of words which is likely to be associated from the motion symbol of motion pattern and to be grammatically correct. The search problem can be described as follows.

$$\boldsymbol{\omega^o} \;=\; \arg\max_{\forall\boldsymbol{\omega}} P(\ \boldsymbol{\Omega}^\omega|\lambda^o,M)P(\boldsymbol{\omega}|\ \boldsymbol{\Omega}^\omega,L) \quad (8)$$

where M and L is the motion language model and the natual language model respectively. $\boldsymbol{\omega}$ is a sequence of words; $\boldsymbol{\omega}=\left\{\omega_1^*\omega_2^*\cdots\omega_{n_*}^*\right\}$. $\boldsymbol{\Omega}^\omega$ is a set of words in the sentence $\boldsymbol{\omega}$. The first term and the second term in Eqn.8 evaluate the likelihood that the motion symbol generates the set of words composing the sentence and the likelihood that the sentence is generated by the set of words. Eqn.8 is solved by using beam search algorithm. Beam search is a best-first search which reduces its resource requirement by keeping a predetermined number of best partial solutions as candidates.

2) Generation of Motion Pattern from Sentence: A motion pattern is generated from a given sentence through searching a motion symbol of motion pattern corresponding to the sentence and regenerating temporal-spatial data from the motion symbol. The search problem for the motion symbol corresponding to the sentence can be described as follows.

$$\begin{aligned}\lambda^o &= \arg\max_{\lambda_i:i=1,2,\cdots,N_\lambda} P(\lambda_i|\boldsymbol{\Omega}^\omega, M)P(\boldsymbol{\Omega}^\omega|\boldsymbol{\omega}, L)\\ &= \arg\max_{\lambda_i:i=1,2,\cdots,N_\lambda} \frac{P(\boldsymbol{\Omega}^\omega|\lambda_i, M)P(\lambda_i)}{P(\boldsymbol{\Omega}^\omega)} P(\boldsymbol{\Omega}^\omega|\boldsymbol{\omega}, L)\\ &= \arg\max_{\lambda_i:i=1,2,\cdots,N_\lambda} P(\boldsymbol{\Omega}^\omega|\lambda_i, M)P(\boldsymbol{\Omega}^\omega|\boldsymbol{\omega}, L)\\ &= \arg\max_{\lambda_i:i=1,2,\cdots,N_\lambda} \prod_{j=1}^{n_*} P(\omega_j^*|\lambda_i)P(\boldsymbol{\Omega}^\omega|\boldsymbol{\omega}, L) \qquad (9)\end{aligned}$$

where we use an assumption of equiprobable prior probabilities of motion symbol: $P(\lambda_i) = \dfrac{1}{N_\lambda}$. Eqn.9 is also solved by using beam search algorithm.

III. Experiments

A. Experimental Result of Interpreting Motion Pattern as Sentence

The proposed framework of combining the motion language model and the natural language model was tested on human motion data obtained through a optical motion capture system. The motion capture system measures the positions of 34 markers attached to a performer. The sequences of marker positions are converted to the sequences of joint angles by inverse kinematics computation based on a humanoid robot with 20 degrees of freedom. The human motion data set contains 25 kinds of motion patterns related to baseball such as “running”, “jumping”, “swinging a bat”, “throwing a ball”, and so on. The average length of the motion data is 3.09[sec]. 25 HMMs are optimized by using these motion data and the motion symbols are acquired.

The human motion data can be recognized as one of the HMMs. The data is also expressed by a sentence manually. Pairs of motion symbols and sentences are used as training data for the motion language model. Table.I shows some examples of training data for the motion language model. The correspondence of motion symbols to words are learned as probabilities that words are associated from the motion symbol. Note that the number of hidden states in the motion language model was set to 50.

TABLE I
Some examples of training data for motion language model.

motion synbols	sentences
1	a player runs
	a runner runs
	a hitter a hitter
2	a player shakes a hand
	a runner shakes a hand
	a hitter a shakes a hand
3	a player swings
	a hitter swings
	a hitter swings a bat
4	a pitcher throws
	a player throws
	a pitcher throws a ball
5	a player opens his arms
	a pitcher opens his arms
6	a player applauds
	a pitcher applauds
	a manager applauds
	a coach applauds

The natural language model consists of 217444 nouns, 1796 adjectives, 3024 adverbs, 14719 verbs, 170 conjunctions, 206 prefixes, 1294 suffixes, 189 prepositions, 15 auxiliary verbs, 231 interjections and 430 ohter words, and 2429 word classes. Note that the words unrelated to baseball are also included in the natural language model.

Table.II shows 5 sentences which are generated from each motion symbol, the likelihood that the words in the sentences are associated in the motion language model, the likelihood that the sentence is generated by the natural language model, and the likelihood of the co-occurrence. The motion of “running” is interpreted as some sentences : “a player runs”, “a hitter runs” and “a runner runs” , which are same as training sentences. The motion of “shaking a hand” is also correctly interpreted as sentences : “a player shakes a hand”, ‘a hitter shakes a hand” and ‘a runner shakes a hand”. The motion of “swinging a bat” is incorrectly interpreted as a sentence :“a player a player”, which has the largest likelihood but does not make sense. However the second candidate sentences : “a player swings” is appropriately associated from the motion pattern. The likelihood that “a player a player” in the first sentence is generated by the motion language model is smaller than that in “a player swings”. But “a player swings” is evaluated as grammatically worse sentence than “a player a player”. The motion of “applauding” is interpreted as correct sentences. The correspondence of the motion symbol and the sentences is appropriately acquired. The true-positive rate in all motion symbols in the training set is 84%.

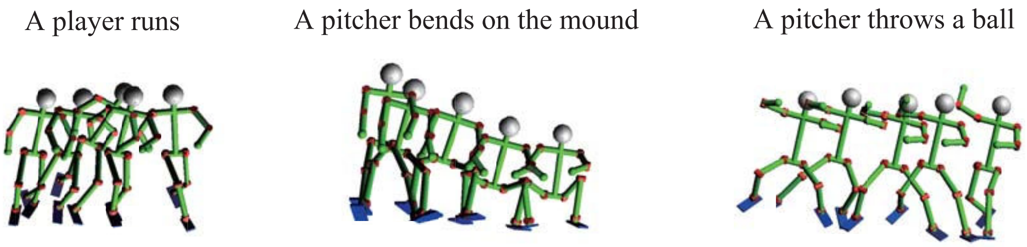
B. Experimental Result of Generation of Motion Pattern from Sentence

The linguistic processing to generate the motion pattern from the sentence was tested. Fig.4 shows the motion data generated from the three kinds of sentences, “subject and verb”, “subject, verb and object”, and “subject, verb and adverb”. The generated motion pattern can be categorized into the motion patterns of “running”,“throwing” and “crouching” respectively. These motion symbols appropriately correspond to input sentences. Therefore, this experiment demonstrates the validity of the framework to generate the motion patterns from not only simple sentence described in “subject and verb” but also a little complicated sentences in “subject, verb and object” and “subject verb and adverb”.

IV. Conclusion

The contributions of this paper are summarized as follows:

1) This paper describes the motion language model which connects words to motion patterns via latent variables



A player runs

A pitcher bends on the mound

A pitcher throws a ball

Fig. 4. A motion symbol is associated from a sentence. The motion symbol generates motion pattern which is represented as a sequence of joint angles. The appropriate motion pattern data is generated from each sentence : "a player runs" and "a pitcher bends on the mound" and "a pitcher throws a ball".

TABLE II

EXPERIMENTAL RESULTS. THIS TABLE SHOWS SENTENCES WHICH MOTION PATTERNS ARE INTERPRETED AS.

motion symbols	sentences	$\log P(\mathbf{\Omega}^{\omega}\|\lambda, M)$	$\log P(\boldsymbol{\omega}\| \mathbf{\Omega}^{\omega}, L)$	$\log P(\mathbf{\Omega}^{\omega}\|\lambda, M) + \log P(\boldsymbol{\omega}\| \mathbf{\Omega}^{\omega}, L)$
1	a player runs	-7.53	-12.74	-20.28
	a hitter runs	-7.53	-14.60	-22.14
	a runner runs	-7.53	-14.87	-22.40
	a player a player	-8.63	-13.81	-22.44
	a player a hitter	-8.63	-15.67	-24.40
2	a player shakes a hand	-7.47	-15.41	-22.88
	a hitter shakes a hand	-7.69	-17.26	-24.96
	a runner shakes a hand	-7.47	-17.53	-25.00
	a player a hitter	-9.35	-15.67	-25.02
	a hitter a player	-9.35	-15.67	-25.02
3	a player a player	-9.13	-13.81	-22.94
	a player swings	-8.03	-16.65	-24.68
	a player a bat	-9.13	-16.06	-25.19
	a bat a player	-9.13	-16.06	-25.19
	swing a player	-9.13	-16.19	-25.32
4	a player throws	-8.03	-13.49	-21.52
	a pitcher throws	-7.34	-14.49	-21.83
	throw a pitcher	-8.44	-13.63	-22.07
	a ball throws	-8.03	-15.03	-23.07
	throw a ball	-9.13	-14.17	-23.30
5	a player opens his arms	-13.17	-21.56	-34.74
	his arms open a player	-13.17	-21.56	-34.74
	a pitcher opens his arm	-13.17	-21.64	-34.82
	his arm, a pitcher opens	-13.17	-21.64	-34.84
	a player a player his arm	-14.97	-21.24	-36.21
6	a player applauds	-7.82	-15.37	-23.20
	a pitcher applauds	-7.82	-16.37	-24.20
	a manager applauds	-7.82	-16.44	-24.27
	a pitcher a pitcher	-9.21	-15.81	-25.02
	a manager a player	-9.21	-15.88	-25.09

stochastically. The motion language model is defined by two kinds of parameters : the probability that the motion pattern generates the latent variable and the probability that the latent variable generates the word.

2) We proposes a novel approach to combining the motion language model with the natural language model. The natural language model was developed for morphological analysis and was optimized by using large-scale of texts. The combining the motion language model and the natural language model makes it possible for robots to have two linguistic processing. One is to interpret motions as sentences and another is to generate motion patterns through association from sentences.
3) These two linguistic processing can be implemented. The combining of the motion language model and the natural language model is confirmed to realized appropriately association between motion symbols and sentences even in the case that the natural language model consists of enormous amount of words.

These two processing is translation between motions and language. We aims at robots that make inferences using a symbolic system of language. The inferences will lead to key technologies of robots assisting humans in various situations and of connecting web mainly based on text analysis to real-world information.

V. ACKNOWLEDGMENTS

This research was supported by Category S of Grant-in-Aid for Scientific Research (20220001), Japan Society for the Promotion of Science.

References

[1] M. Haruno, D. Wolpert, and M. Kawato, "Mosaic model for sensorimotor learning and control," *Neural Computation*, vol. 13, pp. 2201–2220, 2001.

[2] J. Tani and M. Ito, "Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment," *IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans*, vol. 33, no. 4, pp. 481–488, 2003.

[3] H. Kadone and Y. Nakamura, "Symbolic memory for humanoid robots using hierarchical bifurcations of attractors in nonmonotonic neural networks," in *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 2900–2905.

[4] A. Billard and R. Siegwart, "Robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 47, pp. 65–67, 2004.

[5] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *International Journal of Robotics Research*, vol. 23, no. 4, pp. 363–377, 2004.

[6] W. Takano, K. Yamane, T. Sugihara, K. Yamamoto, and Y. Nakamura, "Primitive communication based on motion recognition and generation with hierarchical mimesis model," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2006, pp. 3602–3609.

[7] Y. Sugita and J. Tani, "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes," *Adaptive Behavior*, pp. 33–52, 2005.

[8] T. Ogata, M. Murase, J. Tani, K. Komatani, and H. G. Okuno, "Two-way translation of compound sentences and arm motions by recurrent neural networks," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1858–1863.

[9] P. Chang, M. Han, and Y. Gong, "Extract highlights from baseball game video with hidden markov models," in *Proceedings of the International Conference on Image Processing*, 2002, pp. 609–612.

[10] R. Barzilay and L. Lee, "Catching the drift: Probabilistic content models, with applications to generation and summarization," in *Proceedings of the 2nd Human Language Technology Confenrece and Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, 2004, pp. 113–120.

[11] T. Shibata and S. Kurohashi, "Unsupervised topic identification by integrating linguistic and visual information based on hidden markov models," in *Proceedings of the Joint 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 2006, pp. 755–762.

[12] W. Takano and Y. Nakamura, "Integrating whole body motion primitives and natural language for humanoid robots," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2008, pp. 708–713.

[13] ——, "Statistically integrated semiotics that enables mutual inference between linguistic and behavioral symbols for humanoid robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2009, pp. 646–652.

[14] K. Takeuchi and Y. Matsumoto, "Hmm parameter learning for japanese morphological analyzer," in *Proceedings of the 10th Pacific Asia Conference on Language, Information and Computation*, 1995, pp. 163–172.

[15] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to japanese morphological analysis," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 230–237.

[16] W. Takano and Y. Nakamura, "Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 425–431.

Appendix

A. Optimization of Motion Language Model

The evaluation function can be modified to the following equation by using (2), (3), (4).

$$\Phi = \sum_{k=1}^{N} \log \left[\prod_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j|\lambda^k, \omega_i^k) \frac{P(\omega_i^k, s|\lambda^k, \theta)}{P(s_j|\lambda^k, \omega_i^k)} \right] \quad (10)$$

$$\geq \sum_{k=1}^{N} \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} \log \frac{P(\omega_i^k, s_j|\lambda^k, \theta)}{P(s_j|\lambda^k, \omega_i^k)} \quad (11)$$

$$\equiv \mathcal{F}\left(P(s|\lambda, \omega), \theta\right) \quad (12)$$

Where Jensen's equality is applied to the modification. $\mathcal{F}\left(P(s|\lambda, \omega), \theta\right)$ is the lower limit of the evaluation function.

The lower limit results in the following equation with the help of Bayes rule.

$$\mathcal{F}\left(P(s|\lambda, \omega), \theta\right) =$$
$$\Phi - \sum_{k=1}^{N} \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j|\lambda^k, \omega_i^k) \log \frac{P(s_j|\lambda^k \omega_i^k)}{P(s_j|\lambda^k, \omega_i^k, \theta)} \quad (13)$$

The second term in (13) is Kullback Distance. The maximization of the lower limit $\mathcal{F}\left(P(s|\lambda, \omega), \theta\right)$ under the distributions of the latent state $P(s_j|\lambda^k \omega_i^k)$ is equivalent to the minimization of the Kullback Distance. The Kullback distance has the minimum value of zero if $P(s_j|\lambda^k \omega_i^k)$ is equal to $P(s_j|\lambda^k, \omega_i^k, \theta)$. Therefore, the distributions of the latent variables under the condition that observable data λ^k and ω_i^k are given can be estimated as (1).

In M-step, the model parameters are optimized such that the lower limit $\mathcal{F}\left(P(s|\lambda, \omega), \theta\right)$ becomes the local maximum. The estimation of the set of the model parameters denoted by θ for maximization of the lower limit results in the following calculation.

$$\arg\max_{\theta} \mathcal{F}\left(P(s|\lambda, \omega), \theta\right)$$
$$= \arg\max_{\theta} \sum_{k=1}^{N} \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j|\lambda^k, \omega_i^k) \log P(\omega_i^k|s_j) P(s_j|\lambda^k)$$

Note that $P(s_j|\lambda^k, \omega_i^k)$ can be elimated since it does not depend on the set of model parameters θ. We also utilize the Bayes rule for the modification. We introduce the Lagrange multiplier with the constraint that $\sum_{i=1}^{N_\omega} P(\omega_i|s) = 1$.

$$\mathcal{L} = \sum_{k=1}^{N} \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j|\lambda^k, \omega_i^k) \left[\log P(\omega_i^k|s_j) + \log P(s_j|\lambda^k) \right]$$
$$-\alpha \left(\sum_{i=1}^{N_\omega} P(\omega_i|s) - 1 \right) \quad (14)$$

$$\frac{\partial \mathcal{L}}{\partial P(\omega|s)} = 0 \quad (15)$$

We can solve (15) and then obtain the analytical expression of the optimal model parameter described by (6) . Application the analogical procedure to another kind of model parameter $P(s|\lambda)$ yields the optimal parameter described by (5).