

심층 개념 계층망 기반 가정환경 이동 로봇의 사용자 일상 계획 자동 생성 및 인식 기술

이범진^{1○}, 최진영², 장병탁¹²

¹서울대학교 컴퓨터공학부, 인지과학 협동과정²
{bjlee, jychoi, btzhang}@bi.snu.ac.kr

Home Service Robot Generating and Recognizing User's Plan by Deep Concept Hierarchy

Beom-Jin Lee^{1○}, Jin-Young Choi², Byoung-Tak Zhang¹²

¹School of Computer Science and Engineering, Cognitive Science Program², Seoul National University

요 약

본 논문에서는 가정로봇이 환경을 관찰하여 획득할 수 있는 다양한 정보(사람 위치, 사물, 사람 감정, 행동)를 통합하여 고차원의 인지기능을 모방하는 인지모델을 제안한다. 이러한 정보의 통합을 통해 제안하는 모델은 사람이 하루의 일정을 수행할 때 구별 가능한 기본 단위인 이벤트를 자동으로 생성하고, 순차적인 이벤트를 인식하여 미래 계획까지 예측한다. 기존 계획 인식(Plan Recognition) 분야에서 가정하는 정제된 이벤트 라이브러리를 사용하는 것이 아닌 제안하는 모델은 자동으로 라이브러리를 생성한다. 실험을 통해 게임과 같은 한정된 도메인의 문제뿐만이 아닌 무한한 이벤트가 존재하는 실제 생활의 적용 가능성을 확인한다. 크기가 조정된 터틀봇을 활용하여 실제 가정환경처럼 구성된 실험실에서 로봇이 사람을 항상 따라다니며 실시간의 실험을 수행하였다. 이를 통해 가정로봇이 사람의 일정을 약 90% 구분 가능하였다.

1. 서 론

인공지능의 발전은 과거 전문가 시스템에서부터 자율적으로 사람을 돕는 로봇으로 발전해가고 있다. 따라서 대중들은 로봇이 고차원의 인지기능을 탑재하기를 기대한다. 예를 들면, 감정과 행동을 이해하고 그들의 삶으로부터 지식을 배우고 더 나아가 위험으로부터 보호해주시기를 기대한다. 그에 만족하기 위해서는 로봇이 개념을 쌓아 상황을 인지하고 적합한 행동을 수행할 수 있는 능력이 필수적이다. 최근 정보의 개념화 방법은 [1]에서 제안하는 확률적인 접근 방법으로 연구되고 있으며 [2]에서는 이를 바탕으로 시각-언어에 연관된 개념망 형성 방법이 연구되었다. 또한 미리 정의된 개념을 통해 계획 인식 (Plan Recognition)을 수행하여 상황 및 목표를 이해하고, 계획하는 연구는 [3][4]에서 이루어 졌다.

이를 해결하기 위해 본 논문에서는 지능시스템을 위한 자동 개념망 형성 방법과 그에 따르는 일정 인식 방법이 혼합된 새로운 인지모델을 제안한다. 제안하는 모델은 관찰자가 관찰대상의 의사결정에 영향을 미치지 않는 'key hole' 문제를 다루고 있다.

그림 1에서 우리가 제안하는 전체 시스템을 확인 할 수 있다. 제안하는 시스템은 가정로봇이 사람을 관찰하며 관찰된 정보를 바탕으로 계획 라이브러리를 만들 수 있는 이벤트를 자동으로 생성한다. 그 후 생성된 라이브러리의 관계를 파악하여 계획을 인식하고

미래의 계획을 예측한다. 본 시스템은 2가지로 구성된다. 첫 번째, 관찰대상의 정보 수집 모듈, 두 번째, 자동 계획 라이브러리 생성 및 인식 모듈로 구성된다.

2. 연구 내용 및 방법

2.1 관찰대상 정보 수집 모듈

관찰대상으로부터 정보를 수집하기 위해 아래와 같은 정보 수집 모듈들을 모두 구현하였다.

- SLAM: 로봇 위치 파악 및 지도생성을 위해서 tinySLAM[5]을 사용하였다.
- Human Detect and Follower: 가정로봇이 관측대상인 사람을 실험 시간내에 지속적으로 따라다니기 위해 관측대상만 항상 따라다닐 수 있는 시스템을 YOLO (You Only Look Once: Unified, Real-Time Object Detection) [6]를 사용하여 구현하였다.
- Object Recognizer and Detector: 가정로봇이 관측대상을 항상 따라다니는 중, 설치된 Xtion 카메라로부터 이미지를 서버에 수신 받는다. 서버에서는 수신 받은 이미지를 R-CNN[7]을 통해 물체 감지 및 인식한다.
- Activity Recognizer: 강인한 사람 행동 인식기를 구현하기 위해 새로운 특징점 벡터를 설계하였다. 제안하는 특징점 벡터는 3개의 RGB그리고 깊이 이미지를 사용하여 추출했으며, 시간적 정보를 추출하기 위해 Spatio-Temporal Interest Point (STIP)

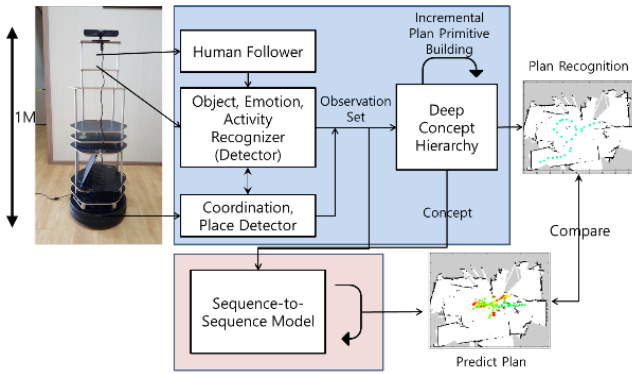


그림 1. 시스템 구조

[8]와 Spatio-Temporal Depth Cuboid Similarity (STDCS) [9] 그리고 순간적 정보를 추출하기 위해 이미지넷을 인식하기 위해 고안된 미리 학습된 컨볼루션 신경망의 최상위 레이어의 특징점을 합친 새로운 특징점 벡터이다. 이를 Softmax로 학습시켰고 구분하려는 행동은 총 9가지 (마시기, 핸드폰사용, 노트북사용, 앉아있기, 걷기, 앉기, 눕기, 일어나기)이다.

2.2 딥 개념 계층망을 이용한 자동 단위 이벤트 생성 및 계획 인식

계획 인식에 필요한 계획 라이브러리 L 은 tuple로 정의된다.

$$L = (A, G, R, P),$$

A 는 유한 집합의 기본 행동, G 는 유한집합의 목표, R 은 목표를 위한 보조 목표 집합, 그리고 P 는 목표를 위한 생성규칙들이다. 전통적 접근법을 사용하여 실생활의 모든 이벤트들을 라이브러리로 직접 정의하는 것은 불가능하다. 그렇기에 본 논문에서는 이 문제를 해결하기 위해 그림 2와 같은 심층 개념 계층망 (Deep Concept Hierarchy)을 사용하여 자동으로 계획 라이브러리를 생성하는 방법을 제안한다.

제안하는 DCH 모델은 [2]에서 최초 제안되었으며 이를 발전시킨 방법을 소개한다. 첫 번째 계층에서는 관찰대상으로부터 받을 수 있는 모든 정보(사람, 사물, 위치, 행동 등)를 입력 받는다. 이렇게 수집된 observation (o) set (그림 2.a)들은 수식 (1)로 이벤트, 즉 계획 라이브러리의 단위(A)가 된다.

$$o \in A \text{ when } V(o) > r \quad (1)$$

$V(o)$ 는 o 가 해당하는 vertex의 값을 나타내며, 한계점 r 은 30으로 정의하였다.

그 후 자동으로 생성된 라이브러리는 상위 계층에서 수식 (2)를 통해 새로운 hyperedge들로 구성이 되고 이는 추후 개념들간의 연관성을 이용하여 새로운 개념들을 창출하는 효과를 가져올 수 있다.

$$\Delta E = \Delta E \cup \{e\} \text{ and } e = \bigcup_{m=1}^k o_m$$

$$P(e) = \prod P(o) \quad (2)$$

ΔE 는 새로운 hyperedge 집합을 나타내며, e 는 생성된 hyperedge를 나타낸다. $P(o)$ 는 o 의 요소들의 총 확률을 나타낸다.

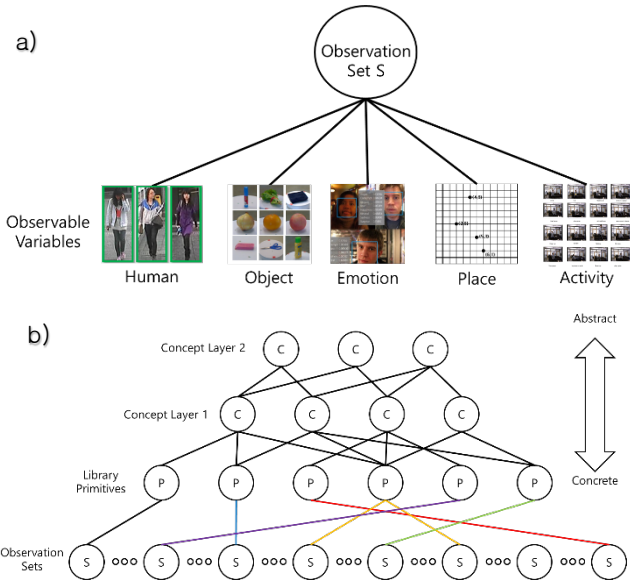


그림 2. 딥 개념 계층망 구조

세 번째 계층은 o 가 등장하는 순서를 고려하여 생성된다. 예를 들어, o_1 이 o_2 보다 먼저 생성이 되었다면 o_1 은 o_2 의 부모가 되는 형식의 e 가 생성되며, 부모 o 는 $i-1$ 번째 o 만 고려한다.

이렇게 학습된 DCH 모델의 세 번째 계층은 라이브러리 단위 간의 연관성을 학습 시켰으며, 이는 곧 계획의 순서를 나타내는 클러스터(R, G) 역할을 한다.

DCH 모델의 학습과정은 그래프 몬테칼로 방법으로 학습이 진행되며 베이지안 규칙을 따르는 수식 (3)의 특징을 따른다.

$$G' = \underset{G_t}{\operatorname{argmax}} P(G_t|O) = \underset{G_t}{\operatorname{argmax}} P(O|G_t)P(G_{t-1}) \quad (3)$$

이는 i 번째 그래프구조는 $i-1$ 번째 구조의 영향을 받으며 충분한 양의 데이터를 관찰하게 된다며 점진적으로 모든 이론 공간을 탐색 할 수 있다. O 는 관측데이터, G_t 는 t 번째 hypergraph이며 $P(G_t|O)$ 를 최대화 하는 방법으로 hyperedge를 생성해 나아가는 방법을 선택한다.

이를 통해 무감독학습방법으로 계획 라이브러리를 생성하고, 연관성을 학습하여 계획을 클러스터링하여 인식 할 수 있는 DCH 모델이 구현된다.

3. 실험 및 결과 토의

실험은 실제 가정환경처럼 구성된 실험실에서 수행되었으며, 3명의 실험자가 주어진 시나리오를 수행하며 진행하였다. 가정로봇은 1M 크기의 ROS가 설치된 core-i7 processor, 8GB memory, GTX960M GPU 노트북을 터틀봇과 연결하여 사용하였고, Ubuntu 14.04가 설치되어있는 12GB Titan X GPU가 설치된 GPU 서버를 사용하였다.

3.1 실험 구성 및 결과

첫 실험은 DCH 모델이 계획 라이브러리를 무감독학습방법으로 생성 가능한지 확인 해 보았다. 수집 데이터의 총 길이는 5160초였으며 실험자는 행동 가능 시나리오

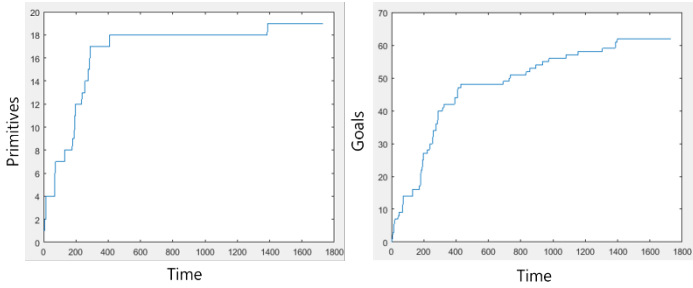


그림 3. 왼쪽-라이브러리 생성 개수, 오른쪽-이벤트간 연관성 형성

오에서 임의로 행동을 선택하여 데이터가 수집되었다. 그 결과 그림 3에서 약 180초에 크기 19의 계획 라이브러리가 생성되었고 계획 이벤트들의 연관성을 그림 4 처럼 확인 할 수 있었다.

라이브러리 생성가능 여부 확인 후 실험자들의 계획 인식이 가능한지 그 여부를 확인하는 두 번째 실험을 수행하였다. 주어진 시나리오는 소파에 앉기→부엌으로 이동 후 식사→소파에가서 노트북하기→소파에 눕기→방에 가서 놀기→거실 걸기→소파에 앉기 순서로 이루어져있다. 해당 시나리오를 각 실험자들이 수행하는 동안 로봇은 항상 따라다니며 실험자를 관찰하였다. 총 데이터 길이는 5461초였고 각 실험자들의 실험 진행시간은 다양했다. 실험 진행 후 실험자들은 자신들의 데이터를 보고 Ground Truth를 만들었고 이를 계획라이브러리로 인식한 결과와 비교하였다. 그 결과 약 89%의 성능을 확인 할 수 있었다. 그림 5는 한명의 실험자에 관한 계획 인식 결과이다.

실험자	A	B	C
관측 데이터 크기	1660	1753	2048
정확도 (%)	91	85	90

표 1. 계획 인식 정확도

3.2 토의

본 논문에서는 실시간으로 가정에서 생활 가능한 로봇이 계획 인식을 수행할 수 있는 새로운 인지모델을 제안하였다. 실험 결과 무감독 학습 방법으로 가정로봇이 관찰대상을 관찰하면서 점진적으로 이벤트를 스스로 생성해 나가고, 그에 대한 연관성을 학습하면서 결과적으로 관찰대상이 하려는 계획을 인식하는 것을 확인 할

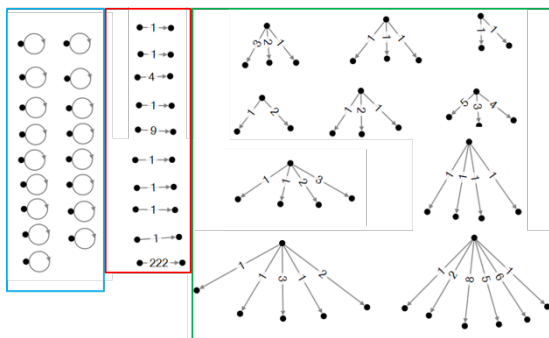


그림 4. 이벤트 간 연관성. 파란색-자기자신의 루프, 빨간색-단일 연관성, 초록색-다중 연관성

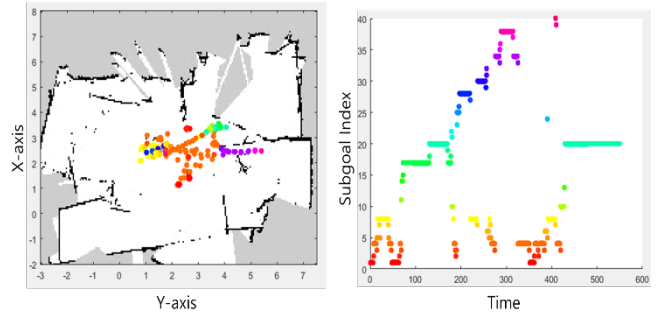


그림 5. 한명의 실험자에 대한 계획 인식 결과, 왼쪽-지도상 표기, 오른쪽-시간 별 이벤트 표기

수 있었다. 추가적으로 인식된 계획 시퀀스를 사용하여 미래를 예측하는 실험을 구상 중이다.

Acknowledgement

이 논문은 2016년도 정부(미래창조과학부, 국방부)의 재원으로 정보통신기술진흥센터(R0126-16-1072-SW스타랩), 한국산업기술평가관리원(10044009-HRI.MESSI, 10060086-RISF), 국방과학연구소(UD130070ID-BMRR)의 지원을 받았음.

Referecne

- [1] B. M. Lake, R. Salakhutdinov, and Joshua B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, 350(6266):1332-8, Dec, 11, 2015.
- [2] J.-W. Ha, K.-M. Kim, and B.-T. Zhang, "Automated construction of visual-linguistic knowledge via concept learning from cartoon videos," In *AAAI*, pp. 522-528. 2015.
- [3] K. Sarah, A. Gal, and E. Karpas. "Goal recognition design for non-optimal agents," In *AAAI*. pp. 3298-3304. 2015.
- [4] F. Kabanza, B. Philipe, B. Francis, R. B. Abder, and I. Hengameh, "Opponent behaviour recognition for real-time strategy games," *Plan, Activity, and Intent Recognition*, 10:05, Jul 7, 2010.
- [5] B. Steux, and O. E. Hamzaoui, "tinySLAM: A SLAM algorithm in less than 200 lines C-language program," In *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on*, pp. 1975-1979. IEEE, 2010.
- [6] J.Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv preprint arXiv:1506.02640*, 2015.
- [7] S. Ren, Ka. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks" In *Advances in Neural Information Processing Systems*, pp. 91-99. 2015.
- [8] I. Laptev, "On space-time interest points," *International Journal of Computer Vision*, 64(2-3), pp.107-123. 2005.
- [9] L. Xia, and J. K. Aggarwal, "Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2834-2841. 2013.