Öğrenilmiş Alt-Hedef Davranışlarının Aktarımı ile Pekiştirmeli Öğrenmenin Robot Navigasyonunda Uygulanması
Implementation of Reinforcement Learning by Transfering Sub-Goal Policies in Robot Navigation

Barış Gökçe, H. Levent Akın
Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
İstanbul, Türkiye
Email: sozbilir,akin@boun.edu.tr

***Özetçe* —Pekiştirmeli öğrenme en yaygın kullanılan yapay zeka yordamlarından biri olmasına karşın boyutluluk sorunundan dolayı istenilen düzeyde kullanılamamaktadır. Durum ve eylem tanım kümelerinin büyümesi durumunda robotun öğrenme hızı dramatik bir şekilde düşmekte ve sonunda robot öğrenemez hale gelmektedir. Boyutluluk sorununu çözmek amacıyla önerilen yöntemler çoğunlukla problemin karmaşıklığını azaltmayı hedeflemektedir. Bazı yöntemlerde karmaşık problem katmanlı bir yapıda modellenirken, diğer yöntemlerde daha basit problemlerde öğrenilen davranışlar doğrudan yeni problemde kullanılmaktadır. Yeni bir problemi en baştan öğrenmek daha önceki tecrübeleri gözardı ederken, eski bilgileri tümüyle aktaran yöntemlerse iki görevin çelişkili gereksinimlerini de aktararak robotu yanlış yönlendirebilmektedir. Bu çalışmadaki temel amaç, robotun öğrenme hızını önceki öğrenme problemlerinde edinilen tecrübelerin ilgili kısımlarını kullanarak arttırmaktır. Önerilen bu yöntemle birlikte problemler katmanlı yapıda modellenilerek bilgi aktarımı sırasında mümkün olduğunca ortak gereksinimler aktarılmaktadır. Önerilen yöntemin performansı player/stage benzetim ortamında robot navigasyon probleminde ölçülmüştür.**

Anahtar Kelimeler—Pekiştirmeli Öğrenme; Hiyerarşik Pekiştirmeli Öğrenme; Bilgi Aktarımı; Robot Navigasyon.

***Abstract*—Although Reinforcement Learning (RL) is one of the most popular learning methods, it suffers from the curse of dimensionality. If the state and action domains of the problem are immense, the learning rate of the agent decreases dramatically and eventually the agent loses the ability to learn. In order to eliminate the effects of the curse of the dimensionality, researchers typically concentrate on the methods that reduce the complexity of the problems. While some of them model the problem in a hierarchical manner, the others try to transfer the knowledge obtained during the learning process of simpler tasks. While learning from scratch ignores the previous experiences, transferring full knowledge may mislead the agent because of the conflicting requirements. The main goal of this study is to improve the learning rate of the agent by transferring the relevant parts of the knowledge acquired as a result of previous experiences. The main contribution of this study is to merge these two approaches to transfer only the relevant knowledge in a setting. The proposed method is tested on a robot navigation task in a simulated room-based environment.**

Keywords—Reinforcement Learning; Hierarchical Reinforcement Learning; Transfer Learning; Robot Navigation.

I. GİRİŞ

Akıllı sistemlerin en önemli öğelerinden birisi öğrenme yeteneğidir. Piaget [1] zekanın gelişimsel yapısı ile ilgili olarak *Kavramsal Gelişim Teorisini* önermiştir. Öğrenme sürecinin gelişimsel yapısı yeni yeteneklerin daha basit eski yetenekleri kullanarak öğrenilmesi şeklinde tanımlanabilir. Örneğin bebekler yürümeden önce yerden kalkmayı öğrenirler. Ayrıca bebeklerdeki öğrenme süreci ilerleyen dönemde koşma olarak devam etmektedir. Benzer bir şekilde robotlar da karmaşık davranışları öğrenebilmek için öncelikle daha basit davranışları öğrenmelidir.

Pekiştirmeli öğrenme akıllı sistemlerde öğrenme yaklaşımı olarak en yaygın kullanılan yöntemlerden birisidir. Hangi durumda hangi hareketin daha iyi bir sonuç getireceğinin tanımlanmasının zor olduğu problemlerde etkin bir şekilde çalışmasına rağmen boyutluluk sorunundan dolayı karmaşık problemlerde öğrenme hızı çok düşmektedir. *Hiyerarşik Pekiştirmeli Öğrenme (HPÖ)* ve *Bilgi Aktarımı* boyutluluk problemini çözmeyi hedefleyen en önemli iki yöntemdir. HPÖ öncelikli olarak karmaşık problemi öğrenmesi daha kolay olan alt-hedeflere bölerek bu alt-hedefleri hiyerarşik bir yapıda tanımlayıp asıl karmaşık problemin çözümüne ulaşır. Bilgi aktarımı yönteminde ise problemin karmaşıklığının eski öğrenme süreçlerinde edinilen bilgilerin aktarımı ile azaltılması hedeflenmektedir.

Bu çalışmanın temel katkısı alt-hedeflerde öğrenilen davranışların bilgi aktarımında kullanılması ile öğrenme sürecinin hızlandırılmasıdır. Önerilen yöntemin temel motivasyonu ise karmaşık problemler ile basit problemler arasında çelişkili gereksinimler olmasının yanında ortak alt-hedeflerin de olmasıdır. Sadece bu ortak gereksinimlerin aktarılması sayesinde işe yarayabilecek eski tecrübeler aktarılabilirken

978-1-4673-5563-6/13/$31.00

çelişkili gereksinimlerin yan etkilerinden de kurtulmuş olunmaktadır.

Bu bildiri şu bölümlerden oluşmaktadır: II. bölümde mevcut HRÖ ve bilgi aktarımı yöntemlerinden bahsedilmektedir. Önerilen yöntem III. bölümde anlatılmaktadır. Yapılan deneyler ve elde edilen sonuçlar ise bölüm IV'da açıklanmaktadır. İleriye yönelik olası araştırma konularından bölüm V'da bahsedilmektedir.

II. İLGİLİ ÇALIŞMALAR

HPÖ yönteminde problemler katmanlı bir yapıda modellenmekte ve alt-hedefler bu hiyerarşik yapıdaki düğümleri belirlemektedir. Bu nedenle alt-hedeflerin kalitesi ve özerk olarak belirlenebilmesi kritik önem taşımaktadır. Özerk alt-hedef belirleme yöntemlerini *çizge temelli* [2], [3], [4], [5], [6] ve *metrik temelli* yöntemler [7], [8], [9], [10], [11] şeklinde iki gruba ayırabiliriz.

Çizge temelli yöntemleri incelediğimizde genel yaklaşım, problemin bir çizge olarak modellenmesi ve çizgenin akışındaki darboğaz özelliği taşıyan düğümlerin belirlenmesi şeklindedir. Bu tip yöntemler çok yavaş çalıştığı için çalışmalar çoğunlukla çizgenin kurulum yöntemleri üzerinde yoğunlaşmaktadır. [12], [13] gibi yöntemler problem tanımından çizgeyi oluştururken, [2], [6] öğrenme sürecinde izlenen gezingeleri kullanmaktadır.

Metrik temelli yöntemler ise her durumun problemin çözümü için ne kadar kritik olduğunu belirlemek için ölçütler tanımlamakta ve bunlara göre derecelendirme yapmaktadır. En yaygın olarak kullanılan metrik robotun her durumdan geçiş sıklığıdır. [7]'de Kretchmar ve diğerleri, başlangıç ve hedef durumlarına yakın bölgelerin problem için darboğaz olmamalarına rağmen ziyaret sıklıklarının çok yüksek olduğuna dikkat çekmişlerdir. Onun için metrik olarak sıklıkla birlikte başlangıç ve hedef durumlarına olan uzaklıkları da dikkate almışlardır. İlgili metrik hesapları Denklem 1, 2 ve 3'te verilmiştir.

$$F_i = \frac{\textit{i. durumu bulunduran gezinimler}}{\textit{Toplam gezinim sayısı}} \tag{1}$$

$$d_i = 2 \cdot \min_{t \in T} \min_{s \in \{s_0, g\}} \frac{|s - i|}{l_t} \tag{2}$$

$$D_i = e^{-1.0 \cdot (\frac{1-d_i}{a})^b} \tag{3}$$

Burada T gezinim kümesini, s_0 başlangıç durumunu, g hedef durumunu, l_t t. gezinimin adım sayısını ifade etmektedir. Karar için kullanılan metrik de $F_i \times D_i$ şeklinde tanımlanmıştır.

Boyutluluk sorununu aşmak için önerilen ikinci yöntem ise *Bilgi Aktarımı*'dır [14], [15], [16], ve [17]. Bu yöntemin temel amacı karmaşık bir hedef verildiği zaman önceki eğitimlerde edinilen bilgileri kullanarak hızlı bir şekilde yeni hedefi öğrenilebilmesidir. Bilgi aktarımı yönteminin tanımında iki görev bulunmaktadır: kaynak, ve hedef görev. Bu görevlerin tanımlandığı kümelere bakarak yöntemi iki ana başlıkta toplayabiliriz: aynı tanım kümesi içinde ve farklı tanım kümeleri arasında. Genel olarak pekiştirmeli öğrenmenin hızını arttırmasına rağmen bu yaklaşımın iki temel problemi bulunmaktadır. Birinci problem farklı tanım kümeleri arasında bir aktarım yapılacağı zaman kaynak ve hedef görev arasındaki fonksiyonun bilinmesi gerekmesidir. Çoğu araştırmacı bu kısmı göz ardı ederek kullanıcının bu fonksiyonu tanımladığını varsaymıştır. İkinci problem ise kaynak ve hedef görevler arasında çelişkili kararlar olması durumunda robotun yanlış yönlendirilmesidir.

III. YÖNTEM

A. Kısmi Bilgi Aktarımı

Bilgi aktarımı Q-değerlerinin ilklendirmelerinde başarılı bir sezgisel yöntem olmasına rağmen hedef görevin kaynak göreve göre çelişkili bir kararı gerektirmesi durumunda robotu yanlış yönlendirebilmektedir. Bu yan etkiden kurtulabilmek için, hedef ve kaynak görevlerin ortak gereksinimlerininin belirlenmesi ve sadece onlar için öğrenilmiş bilgilerin aktarılması gerekmektedir. Bu çalışmada hiyerarşik yapılarda kullanılan alt-hedeflerin bu ortaklıkları tanımlamada ve aktarım sürecinde kullanılması önerilmektedir. Önerilen yöntemde özerk olarak alt-hedeflerin belirlenmesi için sıklık/uzaklık metriğinin [7] geliştirilmiş halini kullanmaktadır. Öğrenme sürecinin artan zorlukta problemleri öğrenme şeklinde olduğunu bildiğimiz için daha önceki süreçlerde bulunmuş olan alt-hedefler yeni alt-hedeflerin bulunmasında kullanılmaktadır. Bu yöntemin ayrıntıları bölüm III-B'da verilmektedir.

Kısmi bilgi aktarımı yönteminde öncelikli olarak daha önce öğrenilmiş olan alt-hedefler listelenerek kullanıcıdan en faydalı olanın seçilmesi istenir. Kullanıcı robota gerekli bilgiyi verdikten sonra alt-hedefi gerçekleştirme becerisi yeni görevin öğrenim sürecinde kullanılmak üzere aktarılır. Yeni gelen karmaşık görev öğrenildikten sonra bu süreçte elde edilen tecrübeler ve daha önceden elde edilmiş olan tüm alt-hedefler kullanılarak *Artımlı Alt-Hedef Belirleme* yöntemi ile yeni bir alt-hedef belirlenir ve robotun bilgi hazinesinde saklanır. Özet olarak *Kısmi Bilgi Aktarımı* yönteminin adımlarını şu şekilde listeleyebiliriz:

1) Q-değerlerinin sıfırlanması
2) Her problem için
   a) Öğrenilmiş alt-hedeflerin listelenmesi ve ortak alt-hedefin kullanıcıdan istenmesi
   b) Robotun ortak alt-hedefe ulaşabilmesi için gerekli Q-değerlerinin aktarılması, diğer değerlerin sıfırlanması
   c) Yeni görev için Q-değerlerinin öğrenilmesi (Q-Öğrenme)
   d) Yeni alt-hedef adaylarının *Artımlı Alt-Hedef Belirleme* yöntemi ile bulunması

B. Artımlı Alt-Hedef Belirme

Çalışmamızda özerk alt-hedef belirleme yöntemi olarak Kretchmar ve diğerlerinin [7] önerdiği sıklık/uzaklık metriğini kullandık. Fakat problemlerimiz ardışık olarak artan zorluk seviyelerinde verildiği ve öğrenilen bilgilerin sonraki aşamalarda kullanılmasını istediğimiz için daha önceki görevlerde bulunan alt-hedef bilgileri de yeni alt-hedeflerin belirlenmesinde kullanıldı. Amacımız mümkün olduğunca durum tanım uzayını kapsamak olduğu için alt-hedeflerin belli bölgelerde yoğunlaşmasını istemiyoruz. Yeni bulunacak alt-hedefleri daha öncekilerden uzak tutmak amacıyla uzaklık metriğine onları da

ekledik. Böylece uzaklık metriği başlangıç, hedef ve diğer alt-hedeflere olan uzaklıkların en düşüğü olarak tanımlandı ve Denklem 2'i Denklem 4 olarak değiştirdik. Verilen denklemde *SG* alt-hedef kümesini ifade etmektedir.

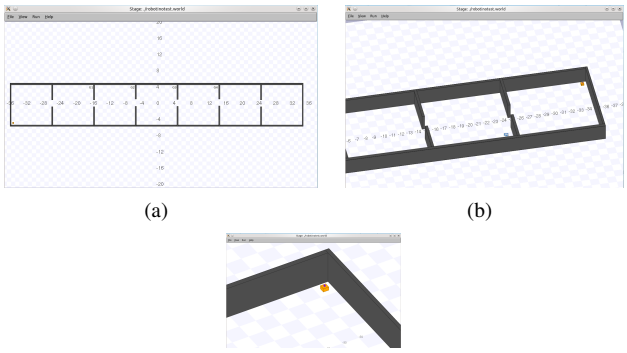
$$d_i = 2 \cdot \min_{t \in T} \min_{s \in \{s_0, g\} \cup SG} \frac{|s - i|}{l_t} \tag{4}$$

Sonuç olarak özerk alt-hedef bulma yöntemini şu şekilde özetleyebiliriz:

1) *sıklık, uzaklık* değerlerinin sıfırlanması
2) Mevcut durumun başlangıç durumu olarak atanması
3) Her başarılı gezinge için
   a) *mevcut durum* == *hedef durum* olana kadar
      i) Mevcut durumun sıklık metriğinin bir arttırılması
      ii) Mevcut durumun başlangıç, bitiş ve alt-hedef durumlarına uzaklığının en küçük olanının hesaplanması
      iii) Mevcut durumun uzaklık metriği olarak uzaklık ve daha önceden atanmış olan değerlerden küçük olanının atanması
      iv) Bir sonraki durumun mevcut durum olarak atanması
   b) Her olası durum için
      i) Mevcut durumunun alt-hedef olma metriğinin sıklık $\times$ uzaklık işlemi ile hesaplanması
4) En yüksek sıklık/uzaklık değerine sahip durum alt-hedef olarak belirlenir

IV. DENEYLER VE SONUÇLARI

Bu çalışmada özellikle boyutluluk sorunuyla ilgilendiğimiz için yaptığımız deneylerde robot zamanla karmaşıklığı giderek artan problemlerle karşılaşmaktadır. Önerilen yöntemin performansını bir robot navigasyon probleminde test ettik. Robotun çalıştığı ve problemlerin tanımlandığı deney ortamı Şekil 1'de gösterilmektedir. Bu ortamda başlangıçta en soldaki odada bulunan robotun sırasıyla Şekil 1.(a)'da görülen G1, G2, G3 ve G4 noktalarına gitmeyi öğrenmesi beklenmektedir. İlk görevde robotun sadece bir kapı geçip yan odadaki bir noktaya gitmesi beklenirken, zamanla geçilmesi gereken kapı sayısı artmakta ve hedef daha da uzaklaşmaktadır.



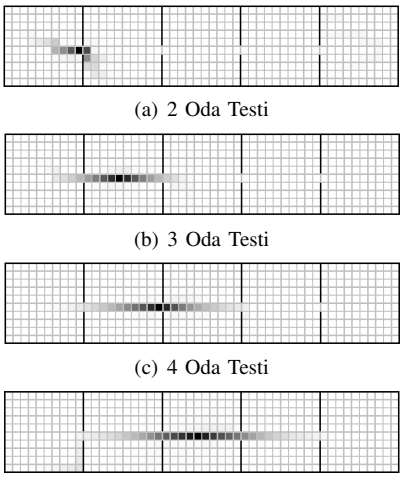
(a) (b)

(c)

Şekil 1. $Player/Stage$ Test Ortamı

Deneylerimizi $Player/Stage$ adlı benzetim ortamında gerçekleştirdik. Robotun durumunu bulunduğu koordinatlarını (x, y) kesikli değeri olarak tanımladık. Ayrıca yapabildiği temel hareketler de kuzeye, güneye, doğuya ve batıya bir birim ilerlemek şeklindedir. Eğer robot geçerli olmayan bir hareket yapmak isterse (duvara doğru hareket etmek gibi) ilerleyemeyeceği için eski yerinde kalmaktadır.

Alt-hedeflerin belirlenmesinde kullanılan sıklık/uzaklık metriğinde yaptığımız değişikliğin performansını incelediğimizde, bulunan alt-hedefleri mümkün olduğunca ortama dağıtmanın faydası olduğu görülmektedir. Şekil 2 özgün yöntem ile bulunan alt-hedefleri gösterirken Şekil 3 bizim geliştirdiğimiz metriğin performansını göstermektedir. Özgün yöntem optimal güzergahın ortalarında öbeklenirken, eski alt-hedeflerin de hesaba katılması ile geliştirilen yöntem alt-hedefleri kapıların yakınlarında bulmaktadır. Bu da bizim daha fazla faydalı bilgi aktarabilmemizi sağlamaktadır.



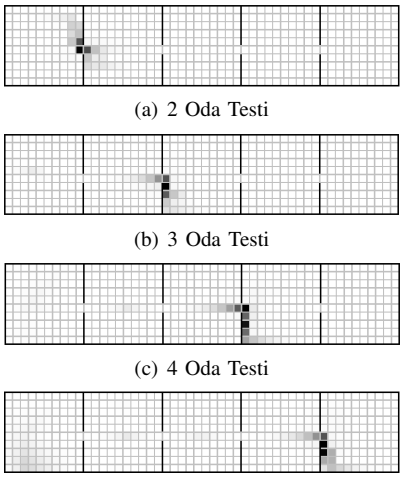
(a) 2 Oda Testi

(b) 3 Oda Testi

(c) 4 Oda Testi

(d) 5 Oda Testi

Şekil 2. Sıklık/Uzaklık Metriği ile Alt-Hedef Belirleme Sonuçları



(a) 2 Oda Testi

(b) 3 Oda Testi

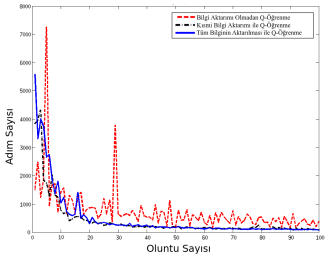
(c) 4 Oda Testi

(d) 5 Oda Testi

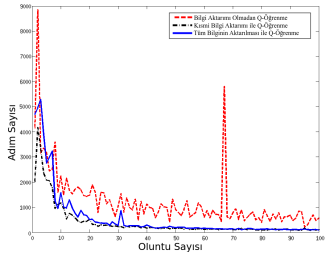
Şekil 3. Artımlı Alt-Hedef Belirme Yöntemi Sonuçları

Önerilen yöntemin performansını hiçbir bilgi aktarımı olmadan ve eski bilgilerin tamamının aktarılması yöntemlerinin performansları ile karşılaştırdık. Öğrenme yöntemi olarak da

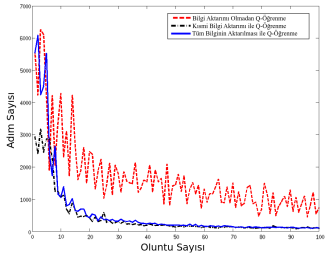
Q-öğrenmeyi kullandık. Geçerli olmayan kararlara (duvara doğru gitmek) -10, geçerli olan kararlara -1 ve hedefe varmasıyla sonuçlanan hareket kararlarına ise +100 ödül verdik. Elde edilen sonuçları Şekil 4'te görmekteyiz. Bilgi aktarılmayan yönteme oranla çok başarılı olmasına rağmen tüm bilginin aktarılması yöntemiyle karşılaştırdığımızda başlangıçtaki denemelerde daha başarılı olduğu görülmektedir. Bunun temel nedeni tüm bilgi aktarıldığında başlangıçta yanlış yönlendirmelerin olmasına rağmen her harekette ceza uygulanmasından dolayı çok geçmeden toparlanmasıdır. Halbuki bizim önerdiğimiz yöntemde robot yanlış yönlendirilmeden farklı olan bölgede arama yapmakta ve hedefi bulması daha kolay olmaktadır.



(a) 3 Oda Testi



(b) 4 Oda Testi



(c) 5 Oda Testi

Şekil 4. Önerilen Yöntemin ve Alternatif Yöntemlerin Performansları

V. SONUÇ

Bu çalışmada yapay öğrenme yöntemlerinin boyutluluk sorununu çözülmesi amaçlanmıştır. Bunun için insanlardaki öğrenme sürecinde olduğu gibi problemleri basitten başlayarak giderek daha zor olacak şekilde tanımladık. Önerdiğimiz yöntemde de verilen görevi öğrenirken aynı zamanda o görev için gerekli alt-hedefleri belirliyor ve ilerleyen zamanlarda zor problemlerle karşılaşıldığında, karmaşıklık seviyesini eski bilgilerimizi kullanarak azaltıyoruz. Performans testlerimiz önerilen yöntemin hiç bilgi aktarılmayan yöntemlere oranla çok daha başarılı olduğunu, tüm bilginin aktarıldığı durumlara göre de başlangıçta daha üstün olduğunu göstermektedir. Bundan sonraki dönemde önerdiğimiz yöntemi gerçek robot üzerinde denemeyi planlamaktayız.

KAYNAKLAR

[1] J. Piaget, *The psychology of intelligence*. London: Routledge & Kegan Paul, 1950.

[2] Ö. Şimşek and A.P. Wolfe and A.G. Barto, "Identifying useful subgoals in reinforcement learning by local graph partitioning," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 816–823.

[3] A. Rad, M. Hasler, and P. Moradi, "Automatic skill acquisition in Reinforcement Learning using connection graph stability centrality," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. IEEE, 2010, pp. 697–700.

[4] P. Moradi, M. E. Shiri, and N. Entezari, "Automatic skill acquisition in reinforcement learning agents using connection bridge centrality," in *Communication and Networking*, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2010, vol. 120, pp. 51–62.

[5] S. Kazemitabar and H. Beigy, "Using strongly connected components as a basis for autonomous skill acquisition in reinforcement learning," in *Advances in Neural Networks – ISNN 2009*, ser. Lecture Notes in Computer Science, W. Yu, H. He, and N. Zhang, Eds. Springer Berlin / Heidelberg, 2009, vol. 5551, pp. 794–803.

[6] N. Negin Entezari, M. Mohammad Ebrahim Shiri, and P. Parham Moradi, "Subgoal discovery in reinforcement learning using local graph clustering," *International Journal of Future Generation Communication and Networking*, vol. 4, no. 3, pp. 13–24, 2011.

[7] R. Kretchmar, T. Feil, and R. Bansal, "Improved automatic discovery of subgoals for options in hierarchical reinforcement learning," *Journal of Computer Science and Technology*, vol. 3, no. 2, pp. 9–14, 2003.

[8] Ö. Şimşek and A. Barto, "Using relative novelty to identify useful temporal abstractions in reinforcement learning," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, vol. 21. Citeseer, 2004, p. 751.

[9] C. Shi, R. Huang, and Z. Shi, "Automatic Discovery of Subgoals in Reinforcement Learning Using Unique-Dreiction Value," in *Cognitive Informatics, 6th IEEE International Conference on*. IEEE, 2007, pp. 480–486.

[10] A. McGovern and A. Barto, "Automatic discovery of subgoals in reinforcement learning using diverse density," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*. Citeseer, 2001, pp. 361–368.

[11] M. Pickett and A. Barto, "PolicyBlocks: An algorithm for creating useful macro-actions in reinforcement learning," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, 2002, pp. 506–513.

[12] B. Hengst, "Discovering hierarchy in reinforcement learning with HEXQ," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*. Citeseer, 2002, pp. 243–250.

[13] I. Menache, S. Mannor, and N. Shimkin, "Q-cut—dynamic discovery of sub-goals in reinforcement learning," *Machine Learning: ECML 2002*, pp. 187–195, 2002.

[14] L. Torrey, J. Shavlik, T. Walker, and R. Maclin, "Relational macros for transfer in reinforcement learning," in *Proceedings of the 17th international conference on Inductive logic programming*. Springer-Verlag, 2007, pp. 254–268.

[15] S. Barrett, M. Taylor, and P. Stone, "Transfer learning for reinforcement learning on a physical robot," in *Ninth International Conference on Autonomous Agents and Multiagent Systems-Adaptive Learning Agents Workshop (AAMAS-ALA)*, 2010.

[16] M. Taylor and P. Stone, "Cross-domain transfer for reinforcement learning," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 879–886.

[17] M. Taylor, P. Stone, and Y. Liu, "Transfer learning via inter-task mappings for temporal difference learning," *Journal of Machine Learning Research*, vol. 8, no. 1, pp. 2125–2167, 2007.