

# Baskın Kümeler ile Hareket Yörüngelerinin Öbeklenmesi

## Clustering Motion Trajectories via Dominant Sets

Çağdaş BAK, Aykut ERDEM, Erkut ERDEM  
Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Ankara, TÜRKİYE  
{n13126810, aykut, erkut}@cs.hacettepe.edu.tr

**Özetçe**—Son yıllarda video gözetleme sistemleri bilgisayarlı görü alanının önemli bir çalışma konusu olarak göze çarpmaktadır. Bu bağlamdaki çalışmalar genellikle video sekanslarında yer alan ortak hareket örüntülerinin belirlenmesi, aykırı hareketlerin tespiti veya olası olayların tahmini üzerinde yoğunlaşmaktadır. Ortaya konan bu çalışmaların başarımları, doğrudan kullanılan ön işleme adımlarının başarımlarına bağlıdır. Bu bildiride, bir video sekansında yer alan hareket yörüngelerini belirleyen ve belirlenmiş olan bu yörüngelerden ortak özellikler gösterenlerinin öbeklenmesine yönelik bir çalışma sunulmuştur. Bu yaklaşım, karmaşık gözetleme sistemlerinin sonuçlarını iyileştirmek için kullanılacak öncül bir adım olarak düşünülebilir. Literatürde yer alan benzer çalışmalardan farklı olarak, önerilen yöntem kullanılan baskın kümeler öbekleme yaklaşımı sayesinde hem ortak hareket örüntülerini daha iyi belirleyebilmekte, hem de uzun süreli video sekansları üzerinde daha başarılı sonuçlar üretebilmektedir. Sunulan algoritmanın etkinliği ve başarımlı VIRAT veri kümesi üzerinden doğrulanmıştır.

**Anahtar Kelimeler** — video işleme; yörünge kümeleme; olay tahmini; gözetleme sistemi; aykırılık tespiti.

**Abstract**— In recent years, video surveillance systems stand out as an important research topic in the field of computer vision. Studies in this context usually focus on detecting common motion patterns in video sequences, determining unexpected motions or predicting possible future events. Performance of these studies directly depends on the performances of the pre-processing steps. In this paper, we present an approach to extract motion trajectories and then to cluster those trajectories that have similar characteristics. The proposed approach can be considered as a preliminary step for complex surveillance systems. Unlike the existing studies in the literature, the proposed method both produces more accurate results for determining common motion patterns and yields successful results on long video sequences in consequence of used clustering algorithm. The effectiveness and the performance of the proposed approach is validated on VIRAT dataset.

**Keywords** — video processing; trajectory clustering; event prediction; surveillance system; anomaly detection.

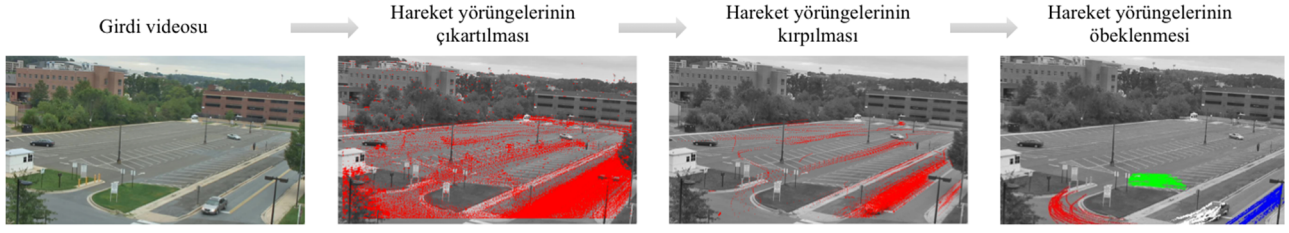


Şekil 1. Hareket halinde ve park etmiş arabalar içeren örnek bir sahne akışı.

### I. GİRİŞ

Video gözetleme ve olay tahmini kapsamında gerçekleştirilen çalışmaların [1-5] çıkış noktaları, insan görsel sisteminin bilgisayar ortamında taklit edilmesine dayanmaktadır. Görsel sistemimiz bir görüntüdeki arka plan, doku, ışık ve nesne etkileşimleri gibi faktörleri kullanarak, görüntüden sahneye ait soyut çıkarımlar yapabilmektedir. Şekil 1’de yer alan örnek sahneyi ele aldığımızda, görsel sistemimiz bu görüntü üzerinden sahnede yer alan arabaların hareket durumları, hareket yönlerini ve hatta hızları ile ilgili soyut çıkarımlar yapabilir. Bu çıkarımları görsel sistemimiz araçların sahnedeki konumları, diğer araçlarla ya da sahne bileşenleriyle olan ilişkileri üzerinden, daha önceki görsel tecrübelerine dayandırarak gerçekleştirir. Sabit görüntüler [3] ya da video sekansları [2] üzerinde çalışan yaklaşımlar benzer bir tecrübeyi bilgisayar sistemlerine denetimli yada denetimsiz öğrenme algoritmaları kullanarak öğretmeyi hedef alırlar.

Bildiride önerilen yöntem, bu bağlamdaki sistemler için bir ön işleme adımı niteliğinde olup, ortak adımlarında gerçekleştirdikleri öğrenme algoritmalarının başarımlarını arttırmaya yöneliktir. Önerilen yöntemde, her bir video sekansında yer alan hareketler Kanade-Lucas-Tomasi [6] algoritması kullanılarak belirlenmiştir. Bildirinin devamında bu hareketlerden yörüngeler (İng. trajectory)



Şekil 2. Önerilen öbekleme yaklaşımının aşamaları.

olarak bahsedilecektir. Takip eden adımlarda ise belirlenen bu yörüngeler, çeşitli ön işlemlerden sonra, önermiş olduğumuz öbekleme yaklaşımı kullanılarak gruplandırılmış ve sonuçlar elde edilmiştir.

## II. ÖNERİLEN YAKLAŞIM

Bildirinin bu aşamasında önerilen yöntem detayları ele alınmakta ve açıklanmaktadır. Önerilen yaklaşım (Şekil 2) temel olarak, girdi video sekansındaki hareketlere ait yörüngelerin belirlenmesi ve bu yörüngelerin anlamlı bir bütünsellik içinde kümelenmesine dayanmaktadır. Elde edilen sonuçlar, gözetleme ya da olay tahmin sistemlerinin sonuçlarını iyileştirecek bir ön işleme adımı olarak kullanılabilir. Önerilen algoritma şu adımlardan oluşmaktadır.

- Hareket yörüngelerinin çıkartılması,
- Hareket yörüngelerinin kırılması,
- Hareket yörüngelerinin kümelenmesi.

### A. Hareket yörüngelerinin çıkartılması

Algoritmanın bu adımında, girdi olarak ele alınan video sekansı ilgili video karelerine (İng. frame) ayrıştırılmıştır. Bu aşamadaki video karesi çıkarım sıklığı (fps), algoritmanın ilerleyen aşamalarındaki sonuçlar üzerinde etkili olmaktadır. Bu çıkarım doğrultusunda, ilerleyen adımlardaki olası hesaplama maliyeti de göz önüne alınarak, ilgili video kareleri saniyede 30 kare sıklığında çıkartılmıştır. Girdi video sekansı ilgili video karelerine ayrıştırıldıktan sonra, video sekansında yer alan hareket yörüngeleri çıkartılır [6]. Algoritmanın bu aşamasında, video sekansı içerisinde yer alan hareketler  $(x, y, t)$  biçiminde temsil edilmiştir. Burada  $(x, y)$  ilgili piksellerin uzamsal alandaki konumlarını belirtirken,  $t$  ise ilgili hareketin zamansal alanda, hangi video karesinde, yani zamanda gerçekleştiğini belirtmektedir. Bu modelleme çerçevesinde, iki ardışık video karesinde belirli bir pencere içerisindeki iki nokta arasındaki farklılığı minimize ederek, ilgili hareketin davranışı belirlenmektedir. Örnekleme gerekirse,  $t$ . zamanda bir piksel  $(x, y)$  konumunda olsun. Aynı piksel  $t + 1$ . zamanda  $(x + n, y + m)$  konumunda ve  $n, m$ 'de belirlenen pencere boyutunun içerisinde ise, ilgili algoritmanın oluşturduğu yörünge listesine, ilgili piksel değerleri için bir satır eklenir. Açıklanan bu algoritmaya

Denklem 1'de yer verilmiştir. Denklem 1, belirli bir pencere içerisinde verilen bir  $p$  noktası için,  $d$  farklılığını minimize eder. Burada,  $I$  ve  $J$  girdi video sekansına ait ardışık video karelerini,  $w(d)$  ise ağırlıklandırma fonksiyonunu temsil etmektedir.

$$\xi(w) = \iint w \left[ J \left( p + \frac{d}{2} \right) - I \left( p - \frac{d}{2} \right) \right] 2w(p) dp$$

$$\xi = \iint w [J(p) - I(p) + g^T(p)d] g(p) w(p) dp = 0 \quad (1)$$

$$g = \left[ \frac{\partial}{\partial x} \left( \frac{I+J}{2} \right) \frac{\partial}{\partial y} \left( \frac{I+J}{2} \right) \right]^T$$

### B. Hareket yörüngelerinin kırılması

Algoritmanın bu adımında, bir önceki adımda çıkartılmış olan hareket yörüngeleri, belirlenen bir eşik değerine göre kırılmaktadır. Muhtemel kamera titremeleri sonucu ortaya çıkan ve istenmeyen gürültüler bu şekilde temizlenmektedir. Bu tarz yörüngelerin sonraki adımlarda performans üzerinde olumsuz etkileri gözlemlenmektedir. Bu adımda ilk olarak her bir hareket yörüngesinin oluşturduğu vektörün uzunluğunu belirlemek için, biricik  $[x, y, t]$  değerleri üzerinden  $x$  ve  $y$ 'ye bağlı olarak standart sapma değerleri hesaplanır. Her bir yörüngeye ait bu standart sapma değeri, önceden tanımlanmış bir eşik değerinden küçük ise gürültü kabul edilir ve veri setinden çıkartılır.

### C. Hareket yörüngelerinin kümelenmesi

Algoritmanın bu adımında ilk olarak yörüngeler arasındaki benzerlik (Denklem 2) tanımlanmaktadır [9]. Bu adımda, bir önceki adımlarda çıkartılmış ve kırılmış olan hareket yörüngeleri, diğer tüm yörüngeler ile karşılaştırılarak bir uzaklık ölçümü elde edilir.

$$dist(P, Q) = \frac{1}{|P|} \sum_{t_p \in P} \min_{t_q \in Q} |t_p - t_q|^2$$

$$w_{ij} = e^{-\frac{dist(P, Q)}{\sigma^2}} \quad (2)$$

Denklem 2'de yer alan benzerlik ölçümünü gerçekleştirebilmek için, her bir yörünge  $(x, y, t)$  formundan,  $T = \{t_i = (x_i, y_i, u_i, v_i)\}_{i=1}^{|T|}$  formuna dönüştürülür. Burada  $x_i, y_i$  ilgili pozisyonu belirtirken,  $u_i$  ve  $v_i$  yörüngesinin hızını belirtir. Yörüngeler için hız değerleri,  $t$  ve  $t + 1$  anlarındaki konum değişimlerine bağlı olarak hesaplanır. Tüm hareket yörüngelerinin birbirleri ile karşılaştırılması sonucunda,  $n$  adet hareket yörüngesi için,  $n \times n$  boyutunda bir diyagonal komşuluk matrisi üretilmektedir.

Girdi video sekansında yer alan hareket yörüngelerine ait komşuluk matrisi tanımlandıktan sonra, ilgili komşuluk matrisi kullanılarak, hareket yörüngeleri anlamsal bir bütünlük içerisinde kümelenmektedir. Literatürde yer alan çalışmalar, bu adımı gerçekleştirmek için genellikle Spektral Öbekleme (SÖ) [8] yöntemlerini tercih etmektedirler. Spektral öbekleme algoritmasının sunmuş olduğu en büyük avantaj, çok yüksek boyutlu matrisler üzerinde bile, hızlı ve etkili çözümler üretebilmesidir. Bildiri de önerilen yöntemde ise literatürde genel olarak kullanılmakta olan yaklaşımın aksine, Baskın Kümeler (BK) öbekleme yönteminin [7] kullanılması önerilmektedir. Bu yöntem grafik teorisi temelli oluşturduğu dominant kümeler üzerinden hiyerarşik bir gruplandırma gerçekleştirmektedir.

İlgili algoritmada kümelenecek olan veri,  $G = (V, E, w)$ , ağırlıklandırılmış çizge modeli olarak tanımlanır. Burada  $V$ , çizgenin düğümlerini,  $E$  ise kenar noktalarını temsil etmektedir. Algoritma bir sonraki adımında elde ettiği bu grafik temsili üzerinden grafikte yer alan kenarlar arasındaki ağırlıklandırmaya bağlı normalize edilmiş bir  $A = (a_{ij})$  diyagonal komşuluk matrisi oluşturur ve öbekleme işlemini bu matris üzerinden gerçekleştirir. Bu adım da ilgili öbekleme algoritmasının oluşturduğu komşuluk matrisi yerine, Denklem 2’de önermiş olduğumuz  $w_{ij}$  komşuluk matrisi kullanılmaktadır. Tüm bunlarla birlikte kullanılan algoritma geleneksel öbekleme algoritmalarından farklı olarak, küme sayısını kullanıcı girdisine bağlı bir parametre olmaktan çıkarmaktadır. Bunun yerine algoritma ilgili girdi komşuluk matrisine bağlı olarak, olası tüm kümeleri üretir. Bu işlemi gerçekleştirirken aynı zamanda üretilmiş olan tüm kümeler için ilgili kümenin ne kadar kaliteli bir küme olduğunu belirten bir skor değeri de üretmektedir. Kullanıcı ilgili skor değeri üzerinden gerçekleştireceği bir eşikleme ile hangi kümeleri tercih edeceğini belirlemektedir.

### III. KULLANILAN VERİ KÜMESİ

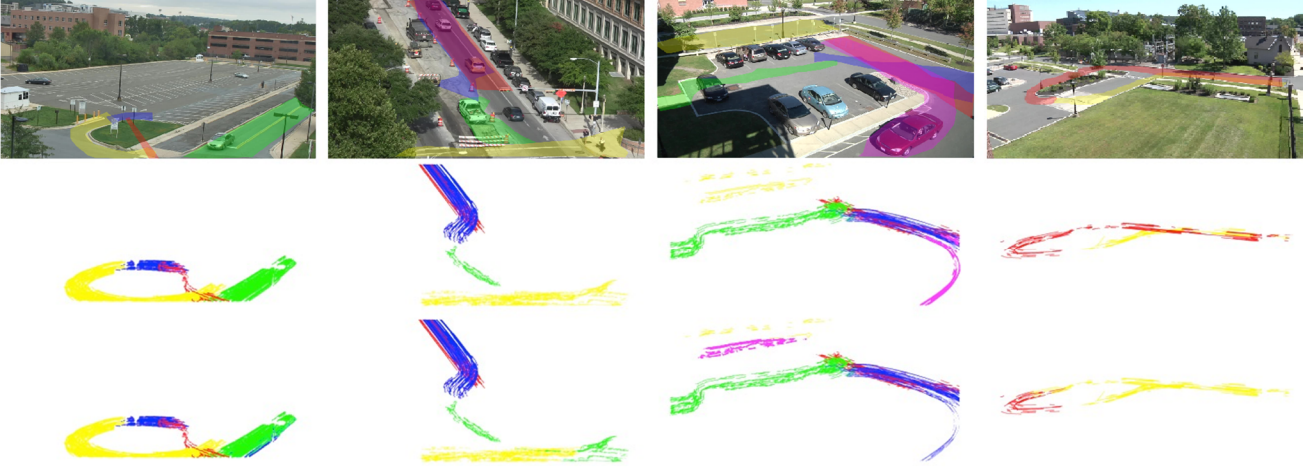
Önerilen algoritmanın etkinliği ve başarımları VIRAT 1.0 veri kümesi [1] kullanılarak doğrulanmıştır. VIRAT veri kümesi bilgisayarlı görü alanında en zorlayıcı veri kümelerinden biri olarak kabul edilmektedir. Hareket tespiti tabanlı bir video veri kümesi olan VIRAT, gerçekçiliği ve doğallığı bakımından benzer video gözetleme veri kümelerinden ayrılmaktadır. VIRAT veri kümesini oluşturan videolar, doğal sahneler üzerinden, herhangi bir kontrollü ortam içermeyen, doğal insan, nesne ve çevre etkileşimlerinden oluşmaktadır. Yüksek çözünürlükte olması ve yüksek video kare sıklığına sahip olması, veri kümesini zor kılmaktadır.

### IV. DENEYSSEL SONUÇLAR

Bildirinin bu bölümünde önerilen algoritmanın başarımları, VIRAT 1.0 veri seti üzerinde gerçekleştirilen deneyler ile görselleştirilmiş ve önerilen öbekleme yönteminin geleneksel yöntemlere göre üstünlükleri belirtilmiştir. Görsel sonuçların yanı sıra kullanılan öbekleme yaklaşımının etkinliği düzenlenmiş RAND indeksi (DRI) (İng. Adjusted Rand Index) [10], normalize karşılıklı bilgi (NKB) (İng. Normalized Mutual Information) [11] ve düzenlenmiş karşılıklı bilgi (DKB) (İng. Adjusted Mutual Information) [12] sayısal metrikleri ile ölçülmüştür. DRI, [1,-1], NKB, [0,1] ve DKB, [0,1] aralığında, yüksek değerlerin daha iyi öbekleme sonuçlarını temsil ettiği sonuçlar üretmektedirler. Bilindiği gibi öbekleme algoritmalarının sayısal başarımlarını ölçmek göreceli olarak hazırlanmış olan mutlak doğruya (İng. ground truth) bağlıdır. Oluşturulan kümelerden hangilerinin doğru, hangilerinin hatalı sonuçlar ürettiği kişiden kişiye yada kullanım amacına bağlı olarak değişebilmektedir. Tüm bu çıkarımlar doğrultusunda deneysel sonuçları üzerinde gerçekleştirilen VIRAT veri kümesinin 4 farklı sahne sekansına ait mutlak doğrular, VIRAT veri kümesi bu amaç doğrultusunda kullanılabilir bir mutlak doğru yada ek açıklama (İng. annotation) içermediği için alan uzmanları tarafından çıkartılmıştır. Her bir hareket yörüngesine ait öbek, tüm yörüngeler üzerinden poligonal olarak temsil edilmiştir. Önerilen algoritmanın, video gözetleme ya da olay tahmini gibi sistemlerin performanslarını iyileştirmeye yönelik bir ön işleme süreci olduğunu düşünecek olursak, ilgili zemin gerçeklikleri de bu amaç doğrultusunda her bir hareketin bütünsel olarak bulunması ve benzer hareketlerin gruplandırılması mantığı ile hazırlanmıştır. Bu doğrultuda hazırlanan zemin gerçeklikleri kullanılarak elde edilen sonuçlara, Tablo 1 ve Şekil 4’te yer verilmiştir. Tablo 1’den de rahatlıkla görülebildiği üzere, önermiş olduğumuz yaklaşımda kullanılan BK, tüm sayısal metrikler baz alındığında geleneksel SÖ kullanan yaklaşımlara üstünlük sağlamaktadır.

	DRI		NKB		DKB	
	SÖ	BK	SÖ	BK	SÖ	BK
Sahne 1	0.37	<b>0.60</b>	0.66	<b>0.77</b>	0.47	<b>0.65</b>
Sahne 2	0.08	<b>0.46</b>	0.41	<b>0.52</b>	0.12	<b>0.53</b>
Sahne 3	0.59	<b>0.75</b>	0.71	<b>0.84</b>	0.57	<b>0.77</b>
Sahne 4	0.36	<b>0.67</b>	0.55	<b>0.77</b>	0.33	<b>0.65</b>
Ort.	0.35	<b>0.62</b>	0.58	<b>0.72</b>	0.37	<b>0.65</b>

Tablo 1. DRI, NKB, DKB karşılaştırmaları.



Şekil 4. Elde edilen bazı öbikleme sonuçları. Üst Sıra: Video karesi ve mutlak doğru, Orta sıra: Spektral öbikleme, Alt sıra: Baskın kümeler

Elde edilmiş olan tüm sayısal sonuçlar, hazırlanmış olan mutlak doğrulardan elde edilmiş ideal küme sayıları kullanılarak üretilmiştir. Örneğin bir video sekansına ait mutlak doğru 4 küme içeriyorsa, SÖ algoritması 4 küme üretecek şekilde çalıştırılmıştır. Şekil 4.'te ise sırasıyla önerilen yaklaşımın kullanıldığı sahne, ilgili sahneye ait BK ve SÖ sonuçları her bir satırda yer almaktadır. SÖ yaklaşımı ile elde edilen sonuçlarda, sahnede yer alan tüm hareket yörüngeleri bulunmaktadır. İlgili video sekansında yer alan hiçbir yörünge, hesaplanan öbekler dışında bırakılmamış ve bir öbeğe atanmıştır. Burada ilgili yöntemin en büyük eksikliklerinden bir tanesi olan, tüm adayları bir kümeye atma isteği ortaya çıkmaktadır. Bu tarz durumlarda küme sayısının kullanıcı girdisine bağlı olarak değişmesinden ötürü, sonuçlar da direk olarak değişmekte ve ilgisiz, hatalı yörüngeler, zorunlu olarak bir sınıfa atanmaktadır. Bunun yanında yine, ilgili algoritma, benzer fiziksel hareketleri gruplamak yerine, sahnede yer alan cisimlerin toplam hareketlerini bir bütün olarak gruplandırmaktadır. Tüm bunlara karşılık önerilen algoritma da kullanılmış olan BK algoritmasının sonuçlarının, olası kullanım alanı oluşturacak sistemler için daha verimli ve kullanışlı sonuçlar ürettiği açıkça ortadadır. Spektral öbiklemin aksine BK, tüm hareket yörüngelerini gruplandırmak yerine, bazılarını öbek dışında da bırakabilmektedir. Bunun yanında, yine Spektral öbikleme yönteminden farklı olarak, algoritma konumsal yakınlığa bağlı hareketleri gruplamak yerine, sadece benzer hareket (yön, yapı) parçalarını gruplamaktadır.

## V. VARGILAR

Bu bildiriye, gözetleme yada olay tahmini sistemleri için bir ön adım olarak kullanılabilir olan bir hareket öbikleme yaklaşımı sunulmuştur. Önerilen algoritmada ilk olarak, video sekansı video karelerine ayrıştırılmış, ayrıştırılan bu karelerden ilgili yörüngeler belirlenmiş ve

son olarak da bu yörüngeler kümelennmişlerdir. Deneysel çalışmalar VIRAT 1.0 veri kümesi üzerinde gerçekleştirilmiştir. Yapılan deneysel çalışmalar sonucunda BK yönteminin ilgili sistemler için uygun ve performansı yükseltici sonuçlar ürettiği görülmüştür. İlerleyen çalışmalarımızda önerdiğimiz bu yaklaşımın, geliştireceğimiz tam bir gözetleme ve olay tahmin sistemine ön adım olması planlanmaktadır.

## VI. TEŞEKKÜR

Bu çalışma kısmen TÜBİTAK 113E497 nolu proje tarafından desteklenmiştir.

## KAYNAKÇA

- [1] S.O. et al. "A Large Scale benchmark dataset for event prediction in surveillance video", CVPR, 2011.
- [2] J.C. Walker, A. Gupta and B. Hebert. "Patch to the future: Unsupervised visual prediction", CVPR, 2014.
- [3] J. Yuen and A. Torralba, Rudnický, "A data driven approach for event prediction", ECCV, 2010.
- [4] Y. Zhu, N. Nayak and A. Roy Chowdhury. "Context-aware modelling and recognition of activities in video", CVPR, 2013.
- [5] Zhong H. et al. "Detection unusual activity in video", CVPR, 2004.
- [6] C. Tomasi and T. Kanade. "Detection and tracking of point features", Rep. CMU-CS-91132, 1991.
- [7] M. Pavan and M. Pelillo. "Dominants sets and hierarchical clustering", IEEE Trans. Pattern Anal. Machine Intell., 2007.29 (1), 167-172.
- [8] Ng. A., Jordan M. and Weiss Y., "On spectral clustering: analysis and algorithm", Advantages in Neural Information Processing Systems, vol 14, pp. 849-856, 2002.
- [9] A. Abrams et. al., LOST: Longterm Observation of Scenes(with Tracks), WACV, 2012.
- [10] L. Hubert and P. Arabie, "Comparing Partitions", 1985, Journal of Classification, 2:193-218.
- [11] Vinh et al., "Information theoretic measures for clustering comparison", ICML, 2009.
- [12] Vinh et al., "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance", The Journal of Machine Learning Research 2837-54, 2010.