

Skor Dağılımlı Üst Arama Modeli

H. Sever, G. Köse
Başkent Üniversitesi Bilgisayar Müh. Bölümü, Ankara
sever@baskent.edu.tr, gkose@baskent.edu.tr

Özetçe

Bir belgenin kullanıcının bilgi ihtiyacı ile ilgililiği hakkında birden çok kanıtın birleştirilmesi, geçmişte ve günümüzde bir çok araştırmaya konu olmuştur. Bu çalışmada farklı arama sistemlerine ait ilgisiz belgelerin skor dağılımlarının eşitlenmesi ile elde edilen ortalama üst arama performansının, literatürde şimdiye kadar rapor edilmiş olan diğer yöntemlerden daha iyi sonuçlar verdiği gösterilmiştir.

Abstract

Combination of multiple evidence of document relevance for effective retrieval has been subject of many past and recent researches. We propose a formal approach to normalizing scores for metasearch by taking the distributions of the scores into account. It is shown that by equalizing the exponential distribution of scores of the top nonrelevant documents yields the best metasearch performance reported in the literature.

1. Giriş

Bilgi erişim sistemlerinde (BES), belirli bir sorgu için belgeler üzerinden bir çok kanıtın elde edilmesi ve bu kanıtların bilgi erişim değerine tekabül edecek şekilde birleştirilmesi problemi üzerine geçmişte ve günümüzde yapılan araştırmalar veya çalışmalar popülerliğini korumaktadır [1, 2, 4, 7, 9, 10, 11, 14, 18]. Bu alanda yapılan çalışmalar, *tekli çerçeve çalışması* (iç birleşim olarak da bilinir) ya da *çoklu çerçeve çalışması* (dış birleşim, veri birleştirme ya da üst arama olarak da bilinir) konularına hedeflenmiştir.

Tekli çerçeve çalışması tek bir sorgu işleme motorunun, bir belge hakkında çoklu kanıtları göz önüne alacak şekilde genişletilmesini hedef

almaktadır. Daha etkin bir erişim¹ için farklı kanıtlar ne olabilir? Bir belge, bilindiği üzere, başlık, özet, vücut, kaynakçalar gibi farklı yapılardan oluşabilir. Bu kısımlar içerisinde geçen anahtar sözcüklerin ilgili belgeyi farklı ağırlıkla temsil etmesi gerektiği deneylerle gösterilmiştir [5]. Hiper-metin veri tabanının dizinlenmesinde, bağlantılar içerisinde geçen sözcüklerin belgeler arası ilgileri çıkarmada nasıl modellenmesi gerektiği son zamanlarda yoğun olarak çalışılan BES konuları içerisinde gelmektedir [6, 12 ,13]. Başka bir bakış açısı, farklı sorgu işleme tekniklerinin tek bir ana sorgu işleme şemsiyesinde farklı ağırlıklarla birleştirilmesidir [1, 2, 16]. Anılan çalışmalarda, çoklu kanıtların doğrusal olarak bir sorgu ifadesinde birleştirilebileceği sonucuna ulaşılmıştır.

Çoklu çerçevede söz konusu olan, bağımsız çalışan birden çok sorgu işleme makinesinin ortak bir amaç kapsamında birleştirilmesidir. Kullanıcı bilgi ihtiyacını karşılamak için aranan belge birden fazla veri tabanında yer alabilir veya birden fazla işleyici ortak veri tabanına aynı bilgi ihtiyacı için erişebilir. Bu durumda, ilgililik kanıt sayısı arttıkça ortalama erişim etkinliğinin önemli bir oranda artacağı beklenmektedir ve bu beklenti deneylerle teyit edilmiştir [2, 17]. Çoklu çerçeve yaklaşımları, Web toplumunda üst arama motorları olarak da anılmaktadır [15]. Günümüzde popüler olan üst arama motorları, aynı bilgi ihtiyacına yanıt veren arama sonuçlarının birleştirilmesi felsefesine dayanmaktadır [7]. Bu tip birleşimler

¹ Bilgi erişim sistemleri, etkinlik çalışmaları yönünden oldukça zengin bir bilim dalıdır (Jones ve Willet, 1997: Bölüm 4). Etkinlik, kısaca, bir sorgu ile ilgili her döndürülen belgenin kullanıcının bilgi ihtiyacını karşıladığı varsayımına, bir BES'in ne kadar yaklaştığını ölçer.

tekli, örtüşen ya da dağıtık veri tabanları üzerinde gerçekleşmektedir. Çoklu çerçeve çalışmasının tek bir veritabanı üzerinde işletilmesi, veri birleştirme; birbirini dışlayan dağıtık veri tabanları üzerinde işletilmesi de koleksiyon birleştirme olarak tanımlanır [17].

Araştırmacılar üst arama problemini skor normalleştirme adımı ve birleştirme adımı olmak üzere iki kısma ayırmışlardır. Skor normalleştirme adımı, erişim sistemleri tarafından üretilen farklı arama sonuçlarında bulunan belgelerin skor değerleri ortak bir ölçüğe getirilerek, skor değerlerinin tüm üst arama sistemi içinde karşılaştırılabilir olmaları sağlanır. Birleştirme adımı ise normalleştirilen skorlar kullanılarak arama sonuçlarının tek bir liste (erişim çıktısı) haline getirilmesi sağlanır. Bu birleşim için çeşitli yöntemler önerilmektedir; bu çalışmada CombSUM (verilen bir sorgu için bir belge ile ilgili döndürülen skorların toplanarak tek bir değer elde edilmesi) ve CombMNZ (CombSUM değerinin sıfırdan farklı skorların sayısı kadar çarpılması) yöntemlerinin her ikisi de deneylerde uygulanmıştır. Bu çalışmada, üst aramada skorların normalleştirilmesi için uygun yöntemin, skorların dağılımlarının göz önünde bulundurulması ile elde edilebileceği önerilmektedir. Verilen bir sorgu için skor dağılımlarının, ilgisiz belgelerin kümesi için üstel bir dağılım ve ilgili belgelerin kümesi için de normal bir dağılım kullanılarak modellenebileceği daha önceki çalışmalarda gösterilmiştir.

Bu çalışmada farklı arama sistemlerine ait ilgisiz belgelerin skor dağılımlarının eşitlenmesi ile elde edilen ortalama üst arama performansının, literatürde şimdiye kadar rapor edilmiş olan diğer yöntemlerden daha iyi sonuçlar verdiği gösterilmiştir. Minimum skorların sıfır olarak belirlenmesi ve farklı üstel dağılımların ortalamalarının eşitlenmesi yöntemi ile ilgisiz dağılımlar düzenlenmiş ve ilgisiz belge skorlarının dağılımını elde etmek için iki farklı yöntem olduğu gösterilmiştir. Bu yöntemlerden birincisinde, bir üstel dağılımın sadece ilgisiz belgeler yerine, ilgili olsun ya da olmasın bütün belgelerin skorlarına uygulanması yöntemi ile ilgisiz belgelerin skor dağılımına yaklaşılmaktadır (Yöntem I). Arama sistemlerinin erişim çıktıları içindeki ilgili belgelerin oranı genellikle küçük olduğu için bu

yaklaşım genel olarak kabul edilebilir. İlgisiz belgelerin dağılımının kestirimi için kullanılan ikinci yöntemde ise, bütün belgelerin skorlarına üstel ve Gaussian modellerinden oluşan karışık bir model bağdaştırılır ve daha sonra da üstel bileşen ilgisiz belgelerin dağılımını kestirmek için kullanılır (Yöntem II). Ele alınan yaklaşımlar TREC -4 ve -9 isimli çok büyük hacimli derlemeler üzerinde test edilmiş ve Yöntem I ve II'nin ortalamaları ile elde edilen normalleştirme değerlerinin CombSUM tekniği ile birlikte en iyi üst arama sonuçları elde edilmiştir.

2. ÜST ARAMA VERİLERİ

Bir üst arama sistemi geliştirilirken performans ölçümlerinin yapılması için en önemli ihtiyaç, geliştirilen sistemlerin üzerinde deneneceği büyük veri kümelerinin bulunması olmuştur. Bilgi erişim sistemlerinin değerlendirilmesinde yöntemler ve kalite testleri yönünden geçmişten gelen oldukça zengin bir birikim vardır. Bilinen test derlemleri CACM, CISI, Cranfield ve NPL olup, tam bilgi verirler [3]; yani, sorgular ve belgeler terim vektörleri cinsinden tanımlı olup, her bir sorgu için ilgili belgeler bir liste halinde tutulur. Bu testler bilgi erişim alanında karşılaşılan meydan okuyucu sorunların çözümü doğrultusunda oluşturulan yeni modellerin test edilmesinde ve daha önemlisi, ortak bazda karşılaştırılmasında zamanla yetersiz kalmışlardır. Bu nedenle, 1990 yılında Amerika İleri Savunma Araştırma Projeleri Ajansı (DARPA), TIPSTER metin projesi çerçevesinde, Ulusal Standartlar ve Teknoloji Enstitüsü'nün (NIST) bilgi erişim teknolojilerini değerlendirmede kullanılmak üzere çok geniş bir metin (ya da genel olarak belge) derlemi oluşturulması istendi.

İlk TREC konferansı 1992 yılında ticari kuruluşların ve çoğu DARPA veya NIST tarafından desteklenen akademik çevrelerin katılımıyla gerçekleştirildiğinde, eldeki derlem 2GB büyüklüğündeki yaklaşık bir milyon belgeden oluşuyordu. Ticari ve akademik bilgi erişim sistemlerinin test yatağı olarak hizmet veren TREC, ulusal kimlikten sıyrılarak zamanla uluslararası bir yarış arenasına haline dönüşmüştür. TREC konferanslarında, katılımcılara çok büyük metin korpusları verilerek, kullanılan erişim tekniklerinin bu veriler üzerinde test edilmesi beklenir. Her bir katılımcı, NIST tarafından

belirlenmiş 50 adet konu için, erişim sistemlerini TREC korpusları(corpora) üzerinde, farklı parametrelerle dört ya da daha fazla defa işletilirler. Erişim sisteminin her bir işletim adımında, verilen konu için elde edilen en üstteki 1000 belge değerlendirmeye katılır. TREC çalışmasının amaçlarından biri de, katılımcılara çok büyük boyutlarda test korpusları sağlamanın yanında kullanılan değerlendirme yöntemlerinde de bir birlikteliğin sağlanmasıdır. Bu çalışmada, TREC-4 ve TREC-9 korpuslarından yararlanılmıştır. Bu korpuslar sırasıyla, TREC-4 olan konular ile adhoc izleri içinde 39 adet işletimi ve TREC-9 için 451-500 arasında olan konular ile web izlerini kapsamaktadır. TREC içinde adhoc izleri, yeni bir konu kullanarak değişmeyen bir belge kümesi içinde arama yapan erişim sistemlerinin performanslarını ölçmek için kullanılmaktadır.

Bu çalışmada aynı zamanda, CLEF (Cross-Language Evaluation Forum) tarafından sağlanan çoklu dil korpuslarından da yararlanılmıştır (<http://clef.iei.pi.cnr.it>) ve birleştirme adımı bulunan bazı sonuçlar TNOU (Twenty - One) erişim sistemi (<http://dis.tpd.tno.n1>) ile karşılaştırılmıştır.

Yeni model ya da tekniklerin denendiği bu konferanslar bir kaç ana görev ve bazıları sonradan ana görev olarak belirlenen pek çok izden oluşmaktadır. Bu bağlamda geliştirilen üst arama tekniklerinin denenmesi ve sonuçlarının değerlendirilmesi amacıyla TREC korpuslarından yoğun olarak yararlanılmaktadır.

3. ÖNCEKİ ÇALIŞMALAR

Bir üst arama motorunda en büyük problem erişim çıktılarının birleştirilerek kullanıcıya tek bir liste olarak sunulması olmuştur. Bu görev gerçekleştirilirken karşılaşılan en önemli problem, farklı arama motorlarının farklı sistemlere sahip olması ve doğal olarak da farklı erişim çıktıları sunmalarıdır. Ancak bir üst arama sisteminin başarıyla gerçekleştirilebilmesi için farklı arama motorlarından elde edilen farklı erişim çıktılarının karşılaştırılabilir bir seviyeye getirilmesi gerekir. Bu problemin çözümü için farklı erişim çıktılarının normalleştirilmesi gereklidir. Normalleştirme işlemi, her birisi kendi ölçeğinde ifade edilen farklı erişim çıktılarının, uygulanan

tekniklerle ortak bir ölçeğe çekilerek karşılaştırılabilir sonuçlara ulaşılması olarak tanımlanabilir.

Farklı erişim çıktıları normalleştirilip karşılaştırılabilir bir ölçeğe çekildikten sonra, erişim çıktılarının birleştirilerek en ilgili belgeler en üstte olacak şekilde kullanıcıya tek bir liste halinde sunulması gerekir. Üst arama motorlarının en kritik görevi olan erişim çıktılarının birleştirilmesinde kullanılmak üzere geliştirilmiş pek çok yöntem bulunmaktadır. Belkin tarafından 1995 yılında yapılan bir çalışmada sunulan farklı birleştirme teknikleri Tablo 1' de sunulmuştur [2];

Tablo1. Belkin tarafından önerilen birleştirme yöntemleri.

Yöntem	İşlev
<i>CombMIN</i>	<i>Bağımsız skor değerlerinin en küçüğü</i>
<i>CombMAX</i>	<i>Bağımsız skor değerlerinin en büyüğü</i>
<i>CombMED</i>	<i>Bağımsız skor değerlerinin ortacası</i>
<i>CombSUM</i>	<i>Bağımsız skor değerlerinin toplamı</i>
<i>CombANZ</i>	<i>CombSUM / sıfırdan farklı skor değerlerinin sayısı</i>
<i>CombMNZ</i>	<i>CombSUM * sıfırdan farklı skor değerlerinin sayısı</i>

CombMIN, CombMAX ve CombMED yöntemleri, erişim çıktılarında bulunan aynı belgeye ait skor değerlerinden tek bir skor değeri seçerken, CombSUM, CombANZ ve CombMNZ yöntemleri ilgili belgenin bağımsız skorlarını kullanarak bu belge için yeni bir skor değeri üretirler. CombMIN yöntemi, ilgili belgeye ait bağımsız skor değerlerinin en küçüğünü alarak ilgisiz bir belgeye yüksek bir sıra (rank) değeri verilme olasılığını en aza indirmeyi amaçlar; CombMAX yöntemi ise düşük olarak derecelendirilmiş olan ilgili belge sayısını azaltmayı amaçlamaktadır. CombSUM, CombANZ ve CombMNZ yöntemlerinin hepsi, bir belge farklı arama motorları tarafından ne kadar fazla sayıda getirilirse, o belgenin ilgili olma olasılığının bir o kadar artacağı varsayımı doğrultusunda geliştirilmişlerdir.

Belkin, bu çalışmasında [2], belge skorlarını en büyük değerli skor değerine bölerek normalleştirdikten sonra TREC verileri üzerinde yaptığı testler sonucunda CombSUM yönteminin bütün yöntemler içinde en iyi birleştirme yöntemi olduğunu göstermiştir. Burada normalleştirme, küresel olarak, birden fazla erişim çıktısı aynı havuzda toplanarak yapılmıştır. Diğer yandan Lee, 1995 yılında yaptığı çalışmada [9], her bir erişim çıktısı kendi içinde normalleştirildiğinde, CombMNZ yönteminin CombSUM yönteminden biraz daha iyi sonuç verdiğini savunmuştur. Erişim sistemlerinin, ilgili belgeler için benzer kümeleri getirmesi ve ilgisiz belgeler için farklı kümeleri getirmesi durumunda en iyi birleşmeyi sağladığı görülmüştür [8,19].

Erişim çıktılarının birleştirilmesi adımında, belgelerin skor değerleri yerine sıra değerlerinin kullanılması tekniği pek çok araştırmacı tarafından incelenmiştir. Belkin tarafından 1995 yılında yapılan çalışmada, birbirleri ile karşılaştırılabilir olmayan skor değerleri üreten farklı erişim sistemlerinin, erişim çıktılarının birleştirilmesinde belge skorları yerine sıra değerlerinin kullanılmasının uygun bir yöntem olduğu savunulmuştur. Lee tarafından 1997 yılında yapılan çalışmada belgelerin skor değerlerini kullanan erişim sistemlerinde sıra değerlerini hesaplamak için aşağıdaki formülün kullanılabileceği gösterilmiştir.

$$SKOR(D) = 1 - (SIRA(D)-1) / R$$

Yukarıdaki eşitlikte $SIRA(D)$, kullanıcı bilgi ihtiyacına karşılık getirilen bir belgenin sıra değerini, $SKOR(D)$, ilgili belgenin skor değerini ve R 'de erişim sistemi tarafından getirilen toplam belge sayısını ifade etmektedir. TREC korpusları üzerinde yapılan testler, erişim çıktılarının birleştirilmesinde, belgelerin sıra değerleri yerine skor değerlerinin kullanılmasının biraz daha fazla erişim etkinliği kazandırdığı gösterilmiştir. Erişim çıktılarının birleştirilmesi konusunda başkaca pek çok araştırma ve yöntem bulunmasına rağmen, bu yöntemlerden hiçbirisi bahsedilen yöntemlerden daha iyi erişim etkinliği sağlayamamıştır.

4. DENEY YÖNTEMLERİ

Bu çalışmada, Montague ve Aslam (2001) tarafından [10], üst arama için önerilen Standart,

Sum ve ZMUV isimli üç farklı normalleştirme yöntemi tartışılmıştır. Bu yöntemler, belge skorlarının doğrusal olarak kaydırılması ve ölçeklenmesi üzerine kurulmuştur (Tablo 2'ye bakınız). Yazarlar, TREC 3-5 ve 9 'u kullanarak, bu üç yöntemin CombSUM ve CombMNZ gibi iyi bilinen birleştirme teknikleri ile birlikte nasıl kullanılacağını göstermişlerdir. $n \in \{2,4,\dots,12\}$ için, n adet arama sistemi sonucunun birleştirilmesinde izlenen yol şu şekilde açıklanabilir: n 'in sabit bir değer olduğu kabul edilerek, bütün arama sistemi sonuçları içerisinden, sonuçların birleşimi için n adet rasgele seçilmiş çıktıdan oluşan bir küme oluşturulur ve bu adım en az 150 deneme için tekrarlanır. Verilen n değeri için performans sonuçları, tüm deneme sonuçlarının ortalamaları alınarak hesaplanır ve rapor edilir [10]. Yaptığımız çalışmada, üst arama sonuçlarının gösterilmesi için bu yöntemin kullanışlı olmadığı tartışılmaktadır. Bu yöntemle elde edilen ortalama performans sonucu, en iyi arama sistemi sonucundan daha düşüktür, bu nedenle, bu sistem için deneylerle belirlenmiş en iyi durumlar şu şekilde ifade edilebilir.

(1) Skorlar, ilgili belgeler için doğru olasılıkları yansıtmalıdır [4, 17];

(2) Dört arama sisteminin erişim çıktıları birleştirildiğinde, üst arama için azalan bir fayda noktasına ulaşılar [17].

Yukarıdaki bilgiler ışığında bizim çalışmamızda aşağıda anlatıldığı gibi, en iyi 5 işletim (her bir işletim bir arama sistemine tekabül etmektedir) göz önünde bulundurulmaktadır.

Tablo 2. Normalleştirme Yöntemleri

Yöntem Adı	Açıklama
Standart	Minimum değeri 0, Maximum değeri 1 olarak belirlir.
Sum	Minimum değeri 0, ve Sum'ı 1 olarak belirlir.
ZMUV	Ortalamayı 0, varyansı 1 olarak belirlir.

TREC-4 (Harman, 1995): Otomatik Adhoc izinden, *Cm1AE*, *pircs1*, *cityo1*, *inq201* ve *siems1* isimli en iyi 5 işletim alınmıştır. Burada verilen bir konu için sorgular sistem tarafından tamamen otomatik olarak üretilirler.

Tablo 3. TREC-4 adhoc izi için, varolan normalleştirme ve birleştirme tekniklerinin ortalama ara değerlendirilmiş kesinlik değerleri.

İşletimler	Sum	ZMUV	Standart	Sum	ZMUV	Standart	Bağımsız İşletim
		CombSUM			CombMNZ		
Cm1AE	0,2944	0,2944	0,2944	0,2944	0,2944	0,2944	0,2944
pircs1	0,3020	0,2948	0,2992	0,3017	0,2938	0,3001	0,2599
cityal	0,3057	0,2933	0,3001	0,3042	0,2923	0,3015	0,2568
INQ201	0,3088	0,2946	0,3056	0,3068	0,2929	0,3052	0,2407
siems1	0,3039	0,2866	0,3046	0,3047	0,2876	0,3059	0,2031
Ortalama	0,3030	0,2927	0,3008	0,3024	0,2922	0,3014	0,251
Değişim(%)	0,1952	0,0201	0,1984	0,2048	0,1641	0,2007	

Tablo 4. TREC-9 web izi için, varolan normalleştirme ve birleştirme tekniklerinin ortalama ara değerlendirilmiş duyarlılık değerleri.

İşletimler	Sum	ZMUV	Standart	Sum	ZMUV	Standart	Bağımsız İşletim
		CombSUM			CombMNZ		
jscbt9w112	0,2801	0,2801	0,2801	0,2801	0,2801	0,2801	0,2801
ric9dpn	0,3017	0,2942	0,2902	0,3008	0,2932	0,2905	0,2616
Nenm	0,3070	0,2977	0,3023	0,3085	0,3014	0,3054	0,2499
acsys9mw0	0,3241	0,3165	0,3261	0,3286	0,3263	0,3310	0,2486
hum9mw0	0,3230	0,3155	0,3227	0,3258	0,3237	0,3257	0,2335
Ortalama	0,3072	0,3008	0,3043	0,3088	0,3049	0,3065	0,2547
Değişim(%)	0,2061	0,1809	0,1947	0,2124	0,1971	0,2034	

TREC-9 (Hawking, 2000): Kısa olmayan kategoriler içinde sunulan işletimler alınmıştır. Bu kategoride, *jscbt9w112*, *ric9dpn*, *Nenm*, *acsys9mwo* ve *hum9tdn* isimli beş işletim alınmıştır.

Bu makale boyunca, performans ölçümleri için ara değerlendirilmiş duyarlılık (non-interpolated precision) değerleri kullanılmıştır. TREC topluluğu tarafından performans ölçümleri için kabul edilen resmi yöntemlerden birisi olan ara değerlendirilmiş duyarlılık, bütün ilgili dokümanların ortalama kesinliğini ifade etmektedir. Bu yöntem, ilgili belgelerin ilgisiz olanların önünde nasıl derecelendirileceğini yansıtan tek değerli bir ölçüyü ifade eder. Biçimsel olarak, bir T konusu için, ara değerlendirilmiş duyarlılık değeri, şu şekilde gösterilir:

$$1 / |Rel | \sum D C Rel | \{D' \in Rel r(D') \leq R(D)\} | r(D) |$$

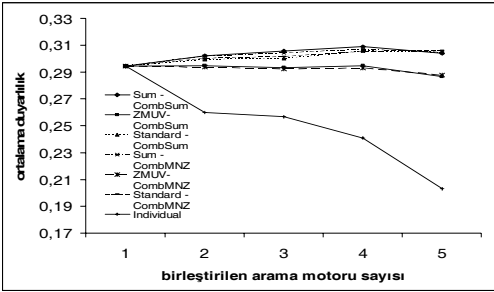
Bu gösterimde $r(D)$, D belgesinin sıra değerini ve Rel , ilgili belge kümesini ifade etmektedir. Bir işletimin 50 konu (veya sorgu) üzerinden performansının hesaplanması için, her bir sorgu

bazında elde edilen etkinlik değerlerinin ortalaması alınır. Bağımsız işletimlere ait ara değerlendirilmiş duyarlılık değerleri Tablo 3 ve 4'de gösterilmektedir.

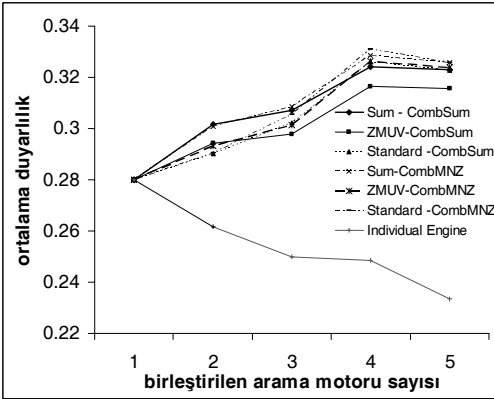
5. VAROLAN NORMALLEŞTİRME YÖNTEMLERİ

Bu çalışmada, *Standart*, *Sum* ve *ZMUV* isimli üç normalleştirme yönteminin, *CombSum* ve *CombMNZ* ile ikili kombinasyonlarının performansları karşılaştırılmıştır. Bu karşılaştırmalar yapılırken, farklı TREC verileri için (TREC -4 ve -9) elde edilen en üstteki 5 işletim ve bunların çıktılarını kullanılmıştır. Performans ölçümlerinden elde edilen sonuçlar Tablo 3 ve 4'de gösterilmektedir. Bu tablolarda ilk satırlar, ortalama ara değerlendirilmiş duyarlılık olarak gösterilen en üstteki işletimin performansını vermektedir. İkinci satırlar, en üstteki iki işletim için birleştirme performansını, üçüncü, dördüncü ve beşinci satırlar ise sırasıyla en üstteki 3, 4 ve 5 işletimlerin erişim çıktılarının birleştirilmesi

performanslarını göstermektedir. İlgili tablolarda birinci sütunlar, veri birleştirme için kullanılan işletim isimlerini, ikinci, üçüncü ve dördüncü sütunlar ise sırasıyla (*Sum, CombSum*), (*ZMUV, CombSum*) ve (*Standart, CombSum*) dan oluşan normalleştirme ve birleştirme çiftlerini göstermektedir. Bu tablolarda son sütunlar ise işleme ilişkin ara değerlenmemiş duyarlılıkları verir ve temel olarak alınır. Tablolarda verilen sonuçlar Şekil 1 ve 2’de grafik olarak da gösterilmiştir.



Şekil 1. TREC-4 adhoc izi için varolan normalleştirme ve birleştirme tekniklerinin ara değerlenmemiş duyarlılıkları.



Şekil 2. TREC-9 web izi için varolan normalleştirme ve birleştirme tekniklerinin ara değerlenmemiş duyarlılıkları.

TREC korpusları temel alınarak üst arama tekniklerine bakıldığında aşağıdaki sonuçlar çıkarılabilir;

- Tablo 3’te görüldüğü gibi, (*Sum,CombSum*) birleşim çifti, %19,52’lik performans değeri ile TREC -4 üzerinde en iyi normalleştirme ve birleştirme yöntem çifti olarak görülmektedir.
- Tablo 4’de görüldüğü gibi, (*Sum,CombMNZ*) birleşim çifti, %21,24’lik performans değeri ile TREC -9 üzerinde en iyi normalleştirme ve birleştirme yöntem çifti olarak görülmektedir.

Her iki durumda da, birleşim tekniği olarak ne kullanılırsa kullanılsın, normalleştirme tekniği olarak Sum’ın kullanılmasının en iyi performans değerine ulaştırdığı sonucuna ulaşabiliriz.

Aslam ve Montaque (2001), tarafından önerilen ZMUV normalleştirme tekniği, iki Gaussian dağılımının varyans ve ortalama değerlerinin normalleştirilmesi olarak görülebilir [10]. Bu kabul, bilgi erişim toplumu için yeni bir yaklaşım değildir. Tablo 5’te gösterilen 2X2’lik olasılık tablosu göz önüne alındığında, aşağıdaki koşullu olasılıkların bu tabloda tutulduğu görülür.

Tablo 5. İlgililik ve erişim için olasılık tablosu. R ve -R sırasıyla ilgili ve ilgisiz belgeleri, Rt ve -Rt sırasıyla, erişilen ve reddedilen belgeleri; a,b,c ve d ilgili hücrelerin sıklığını göstermektedir.

	R	-R	
Rt	a	b	a+b
-Rt	c	d	c+d
	a+c	b+d	a+b+c+d

$a / (a+c) = Pr_R(Rt) =$ hedefi vurmanın koşullu olasılığı ((bir belge geri getirildiğinde (veya arama sistemi ilgili yargısı verdiğinde), gerçekte ilgili olma olasılığı).

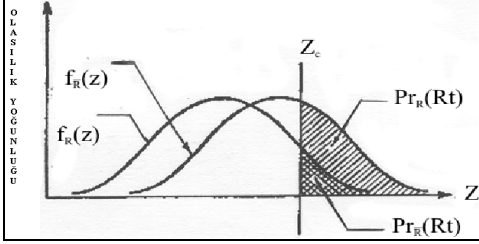
$b / (b+d) = Pr_{-R}(Rt) =$ yanlış alarmın koşullu olasılığı ((bir belge geri getirildiğinde (veya arama sistemi ilgili yargısı verdiğinde), onun gerçekte ilgisiz olma olasılığı).

$c / (a+c) = Pr_R(-Rt) =$ kayıp algılamanın koşullu olasılığı ((bir belge geri getirilmediğinde (veya arama sistemi ilgisiz yargısı verdiğinde), onun gerçekte ilgili olma olasılığı).

$d / (b+d) = Pr_{-R}(-Rt) =$ bir doğru reddetmenin koşullu olasılığı ((bir belge geri getirilmediğinde

(veya arama sistemi ilgisiz yargısı verdiği) onun gerçekte ilgisiz olma olasılığı).

Sweets(1963), olasılık tablosunda yer alan dört sıklık değerinin sıkıştırılarak tek bir etkinlik değeri bulunabileceğini savunmuştur. Buna göre, ilgili ve ilgili olmayan belge skorlarının dağılımları, Şekil 3'te gösterildiği gibi, standart sapmaların birleşimi ile yaklaştırılabilir.



Şekil 3. İlgili ($f_R(z)$) ve ilgisiz ($f_r(z)$) belge skorlarının olasılık dağılımları. Burada, z , belgenin skor değerini ve z_c , eşik değerini göstermektedir [23].

Bu grafikte z_c önceden hata oranını minimuma indireyecek şekilde *hesaplanan* kesme değerini gösterir. Buna göre, eğer $z > z_c$ ise, belge erişim çıktısına yerleştirilir; aksi takdirde arama sistemi belgeyi ilgisiz olarak yorumlar. Bununla birlikte, ilgili ve ilgisiz belge skorlarının dağılımlarının, iki normal dağılımın birleşimi ile modellenemeyeceği yolunda önemli deneysel kanıtlar bulunmaktadır [20, 21, 22]. Bu nedenle, ZMUV tekniğinin etkili sonuçlar vermesi beklenmemektedir. Tablo 3 ve 4'teki sonuçlara bakıldığında da ZMUV-CombMNZ ve ZMUV-CombSum birleşimlerinin en kötü performans değerine sahip oldukları görülmektedir.

Bu noktada, Sum normalleştirme yönteminin neden daha iyi sonuçlar verdiği sorusunun araştırılması gerekmektedir. Monteque ve Aslam tarafından yapılan çalışmada [10], diğer yöntemler arasında bir normalleştirme yönteminin neden seçildiği ile ilgili herhangi bir bilgi verilmemiştir. Ancak yapılan bu çalışmada, uygun normalleştirme yönteminin, farklı arama sistemlerine ait erişim çıktılarındaki ilgisiz belge skor dağılımlarının, normalleştirmede kullanılabilir uygun bir yöntem olduğu ve daha

uygun yaklaşımların nasıl elde edilebileceği gösterilmiştir.

6. DAĞILIM EŞİTLEMESİ YÖNTEMİ İLE BİRLEŞTİRME

Üstel dağılımların olasılık yoğunlukları aşağıdaki eşitlik ile açıklanabilir:

$$P(x) = \lambda \exp(-\lambda x).$$

Bu eşitlikte, rasgele bir D belgesi için $x = \text{score}(D)$ 'yi ifade etmektedir. Üstel dağılımların ortalamasını μ ile gösterirsek, bu değer λ parametresinin tersine eşit olarak kabul edilir:

$$\mu = 1 / \lambda.$$

İlgili belgeler hakkında ön bilgi olmasa bile, ilgisiz dağılımın üstel doğası, belge skorlarının nasıl dağılacığı hakkında bilgi sağlayabilirler. Başka bir deyişle, bu durum, verilen bir sorgu için bir arama sisteminin, döndürülen rasgele bir belge kümesinin skorlarla nasıl eşleştirdiği konusuna açıklık getirmektedir.

Etkili bir normalleştirme yönteminin rasgele seçilmiş belgeleri aynı tarzda düzenlemesi gerektiği için, rasgele dağılımların normalleştirilmesi uygun bir yöntem gibi görünmektedir. Üstel ilgisiz bir dağılım için, standart normalleştirme yöntemleri yani minimum skorlar 0 ve maksimum skorlar 1 olarak ifade edilerek bu işlem gerçekleştirilebilir. Daha sonra ilgisiz bir dağılım için, üstel uygunluk, ML (Maximum Likelihood) yöntemi kullanılarak gerçekleştirilir. Üstel bir dağılım için λ parametresi ML yöntemi ile aşağıdaki gibi kestirilebilir.

$$\lambda = \text{count}(\{D \mid D \in \text{NonRel}\}) / \sum_{D \in \text{NonRel}} \text{SCORE}(D)$$

En son adım, skor dağılımlarının λ parametresini doğrulayan değerler ile eşitlenmesi ile tamamlanır ($\text{score}(D) * = \lambda$). Daha sonra, farklı erişim sistemleri tarafından getirilen erişim çıktıları seçilen birleştirme yöntemi kullanılarak birleştirilir. Bu normalleştirme tekniği EXPML olarak çağrılır. Bu tekniğin uygulanması ile elde edilen sonuçlar, TREC -4 ve -9 korpusları için Tablo 6 ve 7' de verilmiştir.

(EXPML, CombSum) birleşim çiftinin TREC -4 ve -9 korpusları için, diğer yöntemler arasında en iyi sonucu verdiği Şekil 4 ve 5' te görülmektedir. Şekiller, en üstteki 5 işletimin birleşimleri için ortalama ara değerlendirilmemiş duyarlılık değerlerini

Tablo 6. TREC-4 adhoc izi için, bazı normalleştirme ve birleştirme tekniklerinin ortalama ara değerlendirilmemiş duyarlılık değerleri

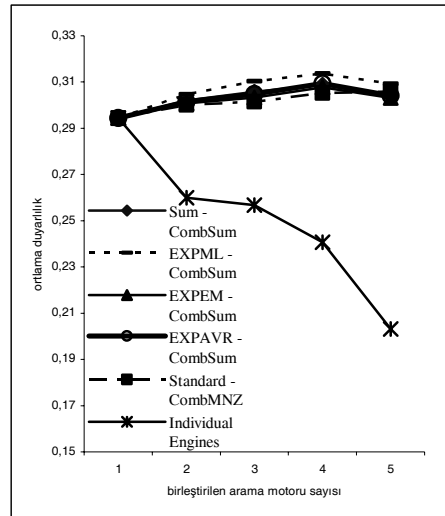
İşletimler	Sum	EXPML	EXPEM	EXPAVR	Standart	Bağımsız
		CombSUM			CombMNZ	İşletim
Cm1AE	0.2944	0.2944	0.2944	0.2944	0.2944	0.2944
pircs1	0.3020	0.3046	0.3008	0.3015	0.3001	0.2599
cityal	0.3057	0.3104	0.3033	0.3048	0.3015	0.2568
INQ201	0.3088	0.3136	0.3077	0.3093	0.3052	0.2407
siems1	0.3039	0.3092	0.3032	0.3041	0.3059	0.2031
Ortalama	0.3030	0.3064	0.3019	0.3028	0.3014	0.2510
Değişim(%)	0.1952	0.2207	0.2028	0.2064	0.2007	

Tablo 7. TREC-9 web izi için ,bazı normalleştirme ve birleştirme tekniklerinin ortalama ara değerlendirilmemiş duyarlılık değerleri.

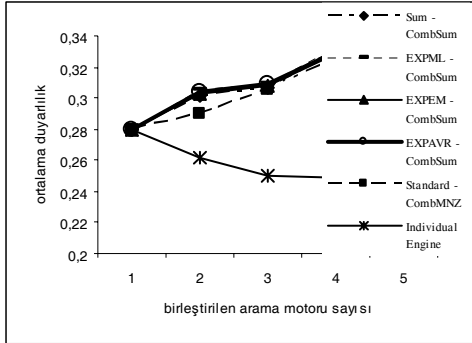
İşletimler	Sum	EXPML	EXPEM	EXPAVR	Standart	Bağımsız
		CombSUM			CombMNZ	İşletim
jscbt9w11	0,2801	0,2801	0,2801	0,2801	0,2801	0,2801
ric9dpn	0,3017	0,3052	0,3030	0,3041	0,2905	0,2616
NEnm	0,3070	0,3096	0,3075	0,3091	0,3054	0,2499
acsys9mw	0,3241	0,3319	0,3281	0,3295	0,3310	0,2486
hum9tdn	0,3230	0,3306	0,3269	0,3288	0,3257	0,2335
Ortalama	0,3072	0,3115	0,3091	0,3103	0,3065	0,2547
Değişim(%)	0,2061	0,2230	0,2136	0,2183	0,2034	

göstermektedir. Her bir birleşim tekniği, normalleştirme yöntemi ve onu takip eden birleştirme tekniğini belirtecek şekilde etiketlenmiştir. Örneğin, CombMNZ yöntemi, standart normalleştirme tekniği ile birlikte kullanıldığında Standart-combMNZ olarak gösterilir. Bu yöntem Lee tarafından [19] en iyi birleştirme tekniği olarak savunulmuştur. EXPML-CombSum yöntemi, TREC -4 ve -9 korpusları için Standart-CombMNZ yöntemine göre en az %1.7 ve %1.6 'lık performans artışı sağlamaktadır.

Pratikte ilgili bilgi bulmak zor olduğu için, ilgili bilgi olmaksızın ilgisiz dağılımların kestirilmesi bir zorunluluk olarak karşımıza çıkar. Bu çalışmada bunu gerçekleştirmek için üç farklı yöntem tartışılmıştır.



Şekil 4. TREC-4 adhoc izi için bazı normalleştirme ve birleştirme tekniklerinin ara değerlendirilmemiş duyarlılıkları.



Şekil 5. TREC-9 web izi için bazı normalleştirme ve birleştirme tekniklerinin ara değerlenmemiş duyarlılıkları.

6.1. KARMA MODEL UYGUNLUĞU KULLANILARAK KESTİRME

Expectation Maximization (EM) yöntemi kullanılarak, üstel ve Gaussian dağılımlarından oluşan karma bir modelin skor dağılımları için uygunlaştırılabileceğini savunulmuştur. Karma bir modelin yoğunluğunu gösteren $p(x)$, aşağıdaki gibi bağımsız bileşenlerin $p(x | j)$ yoğunlukları olarak gösterilebilir.

$$p(x) = \sum_j P(j) p(x|j)$$

Bu eşitlikte j , bağımsız bileşeni; $P(j)$ karma parametreleri gösterir ve

$$\sum_{j=1}^2 P(j) = 1.0 \leq P(j) \leq 1$$

eşitliğini doğrular. Bu çalışmada yoğunluklar, $p(x)$ ve olasılıklar, $P(x)$ ile gösterilecektir. Burada iki farklı bileşenden söz edilmektedir, bunlardan bir tanesi λ ortalama değeri ile birlikte üstel yoğunluk;

$$p(x|1) = \lambda \exp(-\lambda x)$$

ve ortalama μ ve varyans değeri σ^2 ile birlikte bir Gaussian yoğunluğu;

$$p(x|2) = \left(\frac{1}{\sqrt{2\pi\delta}} \right) \exp\left(-\frac{(x-\mu)^2}{2\delta^2}\right)$$

Karma modelin üstel bölümü, ilgisiz dağılımları kestirmek için kullanılırken, Gaussian bölümü ilgili dağılımların bir kestirmesi olur. Verilen bazı

başlangıç değerleri ile parametreleri güncelleyen ve yinelemeli bir teknik olan EM kullanılarak karma model çözülebilir. Bekleme adımında (E-Stop), yürürlükteki değer (üst simge eski ile gösterilir) ve parametrenin yeni değeri (üst simge yeni ile gösterilir) kullanılarak aşağıda gösterilen beklenen değer hesaplanır.

$$E[E^{comp}] = \sum_{n=1}^N \sum_{j=1}^M P^{eski}(j | x^n) \ln \{ P^{yeni}(j) p^{yeni}(x^n | j) \}$$

$$P(j | x) = \frac{p(x | j)P(j)}{p(x)}$$

Bu eşitlikte, E^{comp} bütün veri olasılıklarının beklenen değeri ve $P(j|x)$, verilen x için j bileşenin sonrası (posterior) olasılığını göstermektedir. Maximizasyon adımı (M-step), Gaussian ve üstel bileşenlerin yoğunlukları için belirtilen ifadeler kullanıldıktan sonra beklenen değerlerin maksimumlaştırılmasını gerektirir. Bu adım, parametrelere göre beklenen değerlerin ayrıştırılması ve türevinin sıfır yapılması ile gerçekleştirilir. Bu işlem, yoğunluk ve karışım parametreleri için güncelleme yordamları gerektirir. Karışım parametreleri, aşağıdaki eşitlikler kullanılarak çözülebilir;

$$\mu^{yeni} = \frac{\sum_n P^{eski}(2 | x^n) x^n}{\sum_n P^{eski}(2 | x^n)}$$

$$(\delta^{yeni})^2 = \frac{\sum_n P^{eski}(2 | x^n) \|x^n - \mu^{new}\|^2}{\sum_n P^{eski}(2 | x^n)}$$

$$\lambda^{yeni} = \frac{\sum_n P^{eski}(1 | x^n)}{\sum_n P^{eski}(2 | x^n) x^n}$$

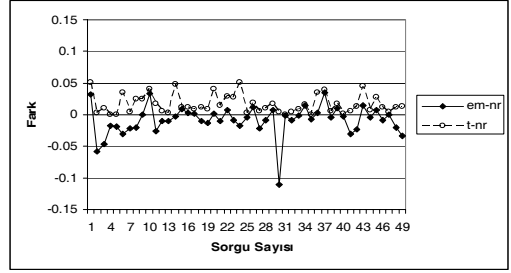
$$P(1)^{yeni} = \frac{1}{N} \sum_n P^{eski}(1 | x)$$

Bu eşitliklerde, güncellenmiş Gaussian parametreleri μ^{yeni} ve δ^{yeni} iken güncellenmiş üstel parametre λ^{yeni} ile verilmiştir. Üstel (P(2)=1-P(1)) için $P(1)^{yeni}$, karışım parametresinin yeni kestirmesidir. $P^{eski}(j|x^n)j^n$, parametrelerin o anki değerleri kullanılarak, verilen j bileşeni için sonraki beklenen değeri ifade eder. EM yöntemi, karışım parametreleri ve bileşen yoğunluklarının başlangıç kestirmelerine ihtiyaç duyar ve parametrelerin değeri daha önceden tanımlanmış bir eşik değerinin altında olduğu zaman sonlanır.

Burada amacımızı gerçekleştirmek için karışımın üstel bölümü ilgisiz dağılımların bir kestirmesi gibi kullanılmıştır. Farklı arama sistemlerinin çıktılarının ilgisiz dağılımlarının eşitlenmesi işlemi, bunların çıktılarını açıklayan karışımların üstel bileşenlerinin eşitlenmesi ile aynıdır. Yani, bütün belgelerin skor değerleri, üstel bileşenlerin anlamlarının aynı olduğundan emin olmak için yeniden ölçeklendirilir. Pratikte, bir arama sistemi üzerinde verilen bir sorgu için her bir belgenin skor değerini üstel bileşenin ortalamasına bölerek bu adımı gerçekleştirmek mümkündür. (EXPEM, CombSum) birleşim çifti için TREC -4 ve -9 korpusları üzerinde elde edilen sonuçlar Şekil 4'te gösterilmiştir. EM algoritması kullanılarak elde edilen üstel dağılımlar, değişik sebeplerden dolayı ilgisiz dağılımlara tam olarak yakınsayamaz.

6.2. TOPLAM DAĞILIM KULLANILARAK KESTİRME

İlgisiz belgelerin dağılımlarını elde etmek için tüm belgelerin dağılımlarının kullanılması genellikle uygun bir yaklaşımdır, çünkü pek çok durumda ulaşılan ilgili belgelerin oranı küçüktür (burada 1000 içinde 34). Şekil 6, CRNLAE sistemindeki 50 sorgu için iki dağılımın ortalamalarının farkını göstermektedir. Bu fark daima pozitif değerdedir ve hepsi için olmasa da pek çok sorgu için düşük bir değerdedir. Şekil 6, aynı zamanda karışık modelden elde edilen üstel modelin ortalamaları ve ilgisiz belge skorlarından elde edilen ortalamalar arasındaki farkı da göstermektedir (em-nr olarak etiketlenmiştir).



Şekil 6. İki yaklaşım tekniği kullanılarak kestirilen ortalamalar arasındaki farklar ve CRNLAE arama sisteminin TREC-4 adhoc izindeki işletimi için ilgisiz dağılımların ortalaması. em, nr ve t sırasıyla, karma modelden elde edilen ortalamalar, ilgisiz dağılımlar ve toplam dağılımları göstermektedir.

Bütün belge skorlarının dağılımları kullanılarak yapılan normalleştirme işlemi, minimum değerlerin 0 olarak belirlenmesi ve tüm belge skorlarının eşitlenmesini ifade etmektedir. Bu sebeple, bahsedilen normalleştirme yönteminin, belge sıra değerleri açısından Montaque ve Aslam tarafından önerilen [10] Sum normalleştirme yöntemine eşdeğer olduğunu söylemek yanlış olmaz (Tablo 2'ye bakınız).

Şekil 4 ve 5'te Sum-CombSum olarak etiketlenmiş olan şekiller, dördüncü veya beşinci en iyi performansa sahiptir ancak, EXPML-CombSum yöntemiyle elde edilen sonuçlara oldukça yakındır. Gerçekten de Sum-CombSum ve EXPML-CombSum yöntemleri arasındaki performans farkı %1.04 -1.39 arasındadır. Bazı veri kümeleri için, Sum-CombSum yönteminin EXPML-CombSum yönteminden daha iyi sonuçlar verdiği Şekil 4'te gösterilmiştir. Bu şekiller TREC -4 ve -9 için en iyi beş işletim sonucunun birleştirilmesi için ortalama kesinlik grafiklerini göstermektedir.

6.3. ORTALAMA KULLANILARAK KESTİRME

Şekil 6, ilgisiz belge dağılımlarının iki farklı kestirmesini göstermektedir. Bu iki kestirmenin ortalamaları alınarak yeni bir kestirme daha elde etmek mümkündür. Normalleştirme ve birleştirme işlemleri TREC -4 ve -9 korpusları için Şekil 4 ve 5'teki şekilleri üretmektedir. Bu yöntem, bütün veri kümeleri üzerinde en iyi performansı sağlamaktadır. Şekil 5'te TREC-9'un web izleri

için, EXPML kullanılarak en iyi sonuçlara ulaşıldığı gösterilmektedir.

6.4. SONUÇLARIN TARTIŞILMASI

Bu çalışmada tartışılan, farklı arama sistemlerine ait ilgisiz skor dağılımlarının eşitlenmesi ve sonuç skorlarının ortalamalarının alınması yöntemi, üst arama için şimdiye kadar rapor edilen en iyi sonuçları üretmiştir. İlgisiz dağılımların kestirilmesi için üç farklı yöntem üzerinde durulmuş ancak en kolay tekniğin toplam dağılımların kestirilmesi (Sum normalleştirme ile aynıdır) olduğu görülmüştür. Sonuçlar, standart adhoc normalleştirme ve birleştirme tekniklerinden bazıları kullanılarak karşılaştırılmıştır. Belge skorlarının olasılıklar olarak ifade edilmesi daha sonrada aşağıda anlatılan yöntemde olduğu gibi bu olasılıkların birleştirilmesi ile sonuçların karşılaştırılması oldukça ilginç görünmektedir.

Verilen bir belge skorunun, sonraki olasılığı hesaplamak için karma model kullanılabilir. Daha sonra farklı arama sistemleri olasılıklarının ortalamaları alınarak birleştirilir. Pratikte sonraki olasılıkların ortalamalarının alınarak birleştirme işleminin yapılması, bu çalışmada önerilen ilgisiz dağılımların eşitlenmesi yöntemi kadar iyi sonuçlar vermemektedir. Bunun sebebi olarak, karma modelin kestirilmesinde oluşan hatalar görülmektedir. Daha önce gösterildiği gibi, Gaussian ve üstel dağılımların kestirilmesi esnasında hatalar oluşmaktadır. Gaussian bileşenlerin kestirilmesi daha zordur çünkü ilişkili belgelerin sayısı küçük olmaktadır.

Çok dillilik ve dağılımlılık durumu düşünüldüğünde, bazı veritabanları için Gaussian bileşenlerinin kestirilmesi çok zor gibi görünmektedir. Örneğin, Türkçe ve İngilizce haber veritabanlarının bulunduğu bir sistem düşünelim. Ankara belediye başkanı hakkında bir sorgu bu sisteme verildiğinde, Türkçe veritabanında belli sayıda ilgili belgeye ulaşılırken, İngilizce veritabanında hiçbir ilgili belgeye ulaşılamayabilir. Böyle bir durumda, sonraki olasılıkların kestirilmesi çok zor görünmektedir. Oysa, bu çalışmada anlatılan yöntem kullanıldığında, karma modelin üstel bileşenleri kabul edilebilir seviyede kestirebilirler (Tablo 8'e bakınız). Burada basit ama ilginç bir biçimde güçlü olan Sum-CombSum

yönteminin özel veritabanları üzerinde nasıl gerçekleştirildiği sunulmaktadır. Bu amaçla CLEF'00 üzerindeki 21 sistemden elde edilen sonuçlar kullanılmıştır. Çok dilli bir İşletim olan ve Tablo 8'de sunulan "tnoutn1" , sorgu dili olarak Almanca kullanır ve "tnoutne1" ,"tnoutnd1" ," tnoutnf1" , "tnoutni1" isimli ve sırasıyla 0.0734, 0.1009, 0.0625 ve 0.0684 performans değerleri ile 4 adet iki dilli orta seviye işletim oluşturulmuştur. Benzer şekilde, "tnoutexsys" isimli çok dilli işletim, sorgu dili olarak İngilizce kullanır ve "en-en" ,"en-de" ,"en-fr" ve "en-it" isimli ve sırasıyla 0.11222, 0.0937, 0.0719 ve 0.0898 performans değerleri ile 4 adet iki dilli orta seviye işletim oluşturulmuştur. "tnoutex1" isimli çok dilli işletimde sorgu dili olarak İngilizce kullanır ve "tnoutee2" , "tnouted1" ,"tnoutef1" ve "tnoutei1" isimli ve sırasıyla 0.0991, 0.0798, 0.0528 ve 0.0765 performans değerleri ile 4 adet iki dilli orta seviye işletim oluşturulmuştur.

Tablo 8, SumCombSum birleşim çiftinin ikili performans değerlerini ve orta seviye işletimlerdeki koleksiyon birleşim işletimlerini göstermektedir. Bu çalışmada "tnout" sistemi ile birlikte Sum-CombSum yönteminin karşılaştırılabilir bir performansa sahip olduğu görülmektedir.

Bu aşamada Gaussian (ilgili) dağılımların eşitlenmesi yönteminin çalışıp çalışmayacağı sorulabilir. İlgili dağılımların aynı parametrelere sahip olacakları (aynı ailenin dağılımları) yönünde herhangi bir bilgi bulunmamaktadır. Fakat pratikte istenen sonuçların elde edilemeyeceği düşünülmektedir.

En iyi bağımsız arama sistemleri üzerinde gerçekleştirilen, çoklu arama sistemlerinin birleştirilmesi genellikle önemli bir miktarda ilerleme sağlamaktadır. Bu bağlamda, en üstteki birkaç işletimin alınması ciddi bir performans artışı sağladığı halde 5' den fazla işletim kullanılması halinde performans artışının durduğu hatta bir kötüleşmeye sebep olabileceği görülmektedir. Bu noktada ilginç olan soru, artık performans artışının görülmediği noktadan önce kaç tane arama makinesinin sonuçlarının birleştirileceğidir.

Tablo 8. CLEF'00 üzerindeki kolektif birleştirmelerin R-P tablosu.

	SUM	WSUM	SUM	WSUM	SUM	WSUM
	CombSUM	tnoutext	CombSUM	tnoutext	CombSUM	tnoutextsys
0	0.5986	0.6945	0.5916	0.6301	0.6894	0.8031
0.1	0.4326	0.4827	0.4054	0.4402	0.4276	0.5395
0.2	0.3624	0.3919	0.3333	0.3738	0.3314	0.4488
0.3	0.2974	0.3266	0.2843	0.3188	0.2850	0.3786
0.4	0.2481	0.2491	0.2339	0.2504	0.2466	0.2934
0.5	0.1971	0.2024	0.2089	0.2178	0.1973	0.2291
0.6	0.1345	0.1432	0.1655	0.1775	0.1581	0.1822
0.7	0.0833	0.0908	0.1232	0.1230	0.1139	0.1427
0.8	0.0578	0.0605	0.0904	0.0992	0.0785	0.1074
0.9	0.0285	0.0274	0.0479	0.0490	0.0411	0.0509
1	0.0107	0.0091	0.0111	0.0105	0.0123	0.0143
Ort.P	0.2049	0.2214	0.2093	0.2256	0.2055	0.2655

7. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, bir sorgu için farklı arama motorlarının erişim çıktılarının en üstünde bulunan ilgisiz belgelerin skor dağılımlarını eşitleyen, üst arama için resmi bir yaklaşım açıklanmıştır. İlişkisiz dağılımların kestirilmesi için 3 farklı yöntem tartışılmış ve bu yöntemlerin üst arama için literatürde şimdiye kadar rapor edilmiş olan en iyi performansa sahip oldukları görülmüştür. Çok dilli arama makineleri üretmek için, farklı veritabanları üzerinde işletilen arama motorlarının erişim çıktılarının birleştirilmesi teknikleri bu çalışma taban alınarak gerçekleştirilebilir. Aynı zamanda, skor normalleştirme problemi, konu algılama ve izleme gibi diğer bilgi erişim görevlerinde de önemli olduğu için, bu çalışmadaki tartışmalar bu alanlara yeni bir bakış açısı kazandırabilir.

8.KAYNAKÇA

[1] Belkin, N., Cool, C., Croft, W., Callan, J., The effect of multiple query representation on information retrieval system performance. In: Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Pittsburg, PA, USA, pp. 339-346, 1993.

[2] Belkin, N., Kantor, P., Fox, E., Shaw, J., Combining the evidence of multiple query representations for information retrieval. Information Processing & Management 31 (3), 431-448, 1995.

[3] C. Buckley., Implementation of the SMART information retrieval system. Cornell University, Technical report 35-686, 1985.

[4] Croft, W., Combining approaches to information retrieval. In: Croft, W. (Ed.), Advances in Information Retrieval. Kluwer Academic Publishers, pp. 1-36, 2000.

[5] Croft, W., Combining approaches to information retrieval. In: Croft, W. (Ed.), Advances in Information Retrieval. Kluwer Academic Publishers, pp. 1-36, 2000.

[6] Deogun, J. S., Sever, H., Raghavan, V. V., Structural abstractions of hypertext documents for web-based retrieval. In: Wagner, R. R. (Ed.). Proceedings of Ninth international Workshop on Database and Expert Systems Applications, (in conjunction with DEXA'98). Vienna, Austria, pp. 385 390, Aug. 1998.

[7] Gauch, S., Wang, G., Gomez, M., Profusion: Intelligent fusion from multiple, distributed search engines. Journal of Universal Computer Science 2(9), 637-649, Sep. 1996. http://www.jucs.org/jucs_2_9/profusion_intelligent_fusion_from.

[8] Katzer, J., McGill, M.J., Frakes, W., DasGupta, P. A study of overlap among document representations. Information Technology: Research and Development 1 (4), 261-274, Oct 1982.

- [9] **Lee, J. H.**, Combining multiple evidence from different properties of weighting schemes, in: Fox, E. A. (Ed.). Proceedings of the 18th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Seattle, WA, pp. 180-188, July 1995.
- [10] **Montague, M., Aslam, J.**, Relevance score normalization for metasearch. In: Proceedings of the ACM 10th Annual International Conference on Information and Knowledge Management (CIKM). Atlanta, Georgia pp. 427-433, November 2001.
- [11] **Manmatha, R., Sever, H.**, A Formal Approach to Score Normalization for Metasearch, In: Human Language Technology Conference (HLT'02) San Diego, March 2002.
- [12] **Pinto, D., McCallum, A.**, Wei, X. and Croft, W.B. Table Extraction Using Conditional Random Fields. In: Proceedings of SIGIR '03 Conference, Toronto, Canada, 2003.
- [13] **Rodrigo, A., Botafoga, R., Ehud, R. Shneiderman, B.**, Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. ACMTIS. pp. 142-180, 1992.
- [14] **Saracevic, T., Kantor P.**, A study of information seeking and retrieving. III. searchers, searches, and overlap. Journal of American Society for Information Science 39(3), 197-216, 1998.
- [15] **Tonta, Y., Bitirim, Y. and Sever, H.**, Türkçe Arama Motorlarında Performans Değerlendirme, Damla Yayınevi Ltd., pp. 1-152, 2002.
- [16] **Tumer, K., Ghosh, J.**, Linear and order statistics combiners for pattern classification, In: Sharkey, A. (Ed.), Combining Artificial Neural Networks. Springer-Verlag, pp. 127-162, 1999.
- [17] **Vogt, C., Cottrel, G.**, Fusion via a linear combination of scores. Information Retrieval 1(2-3), 151-173, 1999.
- [18] **H. Sever and M. Tolun.**, Comparison of Normalization Techniques for Metasearch. ADVIS'02, Lecture Notes in Computer Science, Springer Verlag, Vol. 2457, pp. 133-143, 2002.
- [19] **J. H. Lee.**, Analyses of multiple evidence combination. In the Proc. of the 20th Intl. Conf. on Research and Development in Information Retrieval (SIGIR'97), pp. 267-276, 1997.
- [20] **A. Arampatzis and A. van Hameren**, Maximum likelihood estimation for filtering thresholds. In In the Proc. of the 24th ACM SIGIR conf. on Research and Development in Information Retrieval, pp. 285-293, Sept 2001.
- [21] **Y. Zhang and J. Callan.**, Maximum likelihood estimation for filtering thresholds. In the Proc. of the 24th ACM SIGIR conf. on Research and Development in Information Retrieval, pages 294-302, Sept 2001.
- [22] **R. Manmatha, T. Rath, and F. Feng**, Sept 2001. Modeling score distributions for combining the outputs of search engines. In the Proc. of the 24th ACM SIGIR conf. on Research and Development in Information Retrieval, pages 267-275.
- [23] **Sweets, J.**, 1963. Information Retrieval Systems. Science 141,245-250.