TIPTEKNO'24 MEDICAL TECHNOLOGIES CONGRESS
10-12 October 2024
Yasmin Bodrum Resort, Muğla, Türkiye

COLBERT Mimarisi ile İnce Ayar Yapılan Yöntemler İçin Geri Getirme Performansı Analizi
Analysis of Retrieval Performance for Methods Fine-Tuned with ColBERT Architecture

Berkin Alkan
HAVELSAN
Ankara, Turkey
berkinalkan@havelsan.com.tr

Bekir Bilgehan Tekin
HAVELSAN
Ankara, Turkey
stj.bbtekin@havelsan.com.tr

Alper Karamanlıoğlu
HAVELSAN
Ankara, Turkey
alperk@havelsan.com.tr

İsmail KARAKAYA
HAVELSAN
Ankara, Turkey
ikarakaya@havelsan.com.tr

***Özetçe* —Günümüzde sağlık alanındaki büyük veri kümelerinden bilgiye dayalı geri getirme, önemli bir çalışma alanı olarak öne çıkmaktadır. Veri kümelerinin büyümesi ve çeşitlenmesi, insan müdahalesine ihtiyaç duyulmadan çalışan yapay zeka çözümlerinin önemini artırmaktadır. Sağlık uzmanları için kullanıcının sorgularına uygun bilgilerin hızlı ve doğru bir şekilde sunulması, bilgi erişim süreçlerinin başarısını etkilemektedir. Çalışma, kodlayıcı model tabanlı mimari ile sağlık alanına özgü bilgilerin doğru ve etkin bir şekilde geri getirilmesini amaçlamaktadır. Hız ve performans açısından ColBERT (Contextualized Late Interaction over BERT) mimarisi tercih edilmekte ve modeller bu mimariyle ince ayar yapılarak tıp alanına uyarlanmaktadır. İnce ayar yapılmış modellerin başarıları değerlendirilmektedir. Elde edilen bulgular, tıbbi bilgi geri getirme işlemlerinde, zor negatif örnek madenciliği ve hibrit skor ağırlığı tekniklerinin kullanılmasıyla model performansının artırılabileceğini göstermektedir. Çalışmanın, sağlık alanındaki bilgi erişim sistemlerinin hızını ve doğruluğunu artırma potansiyeli sunmasının yanı sıra, bu alandaki gelecekteki araştırmalar için önemli bir referans niteliğinde olduğu düşünülmektedir.**

Anahtar Kelimeler—ColBERT, ince ayar, BERT, geri getirme, çok dilli

***Abstract*—Nowadays, knowledge-based retrieval from large datasets in healthcare is emerging as an important field of study. The growth and diversification of datasets increases the importance of artificial intelligence solutions that work without the need for human intervention. For healthcare professionals, fast and accurate presentation of information relevant to the user's queries affects the success of information retrieval processes. This study aims to retrieve healthcare-specific information accurately and efficiently with a coding model-based architecture. ColBERT (Contextualized Late Interaction over BERT) architecture is preferred for speed and performance, and the models are fine-tuned with this architecture and adapted to the medical domain. The success of the fine-tuned models is evaluated. The findings show that model performance can be improved by using hard negative instance mining and hybrid score weighting techniques in medical information retrieval. In addition to offering the potential to improve the speed and accuracy of health information retrieval systems, the study is considered to be an important reference for future research in this field.**

Keywords—ColBERT, fine-tuning, BERT, retrieval, multilingual

979-8-3315-2981-9/24/$31.00 ©2024 IEEE

I. GİRİŞ

Son yıllarda büyük veri kümelerinin oluşturulması ile bilgi geri getirme önemli bir konu haline gelmiştir. Bilgi geri getirme yöntemlerinde geleneksel olarak BM25 (Best Matching 25) metin bazlı yaklaşımlar yaygın olarak kullanılmaktadır. Ancak, bu yöntemler kelime eşleşmesine dayalı olduğu için, derin anlamsal benzerlikleri yakalamakta yetersiz kalmaktadır. Metin bazlı aramalar yerine yapay zeka tabanlı kodlayıcı modellerin bilgi geri getirme süreçlerinde kullanılması önem kazanmaktadır [1], [2]. 2018 yılında BERT (Bidirectional Encoder Representations from Transformers) [3] modelinin tanıtılmasıyla, metinlerin anlamsal olarak vektörler halinde temsil edilmesi mümkün hale gelmektedir.

BERT tabanlı yöntemlerde, metinler vektör uzayında temsil edilmekte ve gelen istemlerle vektör benzerliği hesaplanarak en uygun sonuçlar getirilmektedir. Bu şekilde veride kelime eşleşmesi aramasına kıyasla derin anlamsal benzerliklere bakıldığından aranan veriler daha isabetli tespit edilmektedir. Bu amaçla farklı çalışmalar [4], [5] ve [6] ortaya konulmaktadır. Ancak bu yöntemlerde modellerin her biri, farklı amaçlar için optimize edilmiş olup bilgi geri getirme süreçlerinde belirli sınırlamalara sahiptir. Bu yöntemler daha çok dil modelleme görevlerinde kullanılırken, bilgi geri getirme görevlerinde etkinlik ve verimlilik açısından yetersiz kalmaktadır. Başka bir örnek olarak Sentence-BERT yöntemi [7], cümle bazında benzerliklerin tespiti için kullanılmaktadır. Model, anlamsal olarak aynı ve farklı cümle çiftleriyle eğitilerek cümle benzerliklerini bulmada başarılı olsa da, büyük veri kümelerinde istem ve doküman gösterimlerinde yeterince detaylı bir etkileşim sunamamaktadır.

ColBERT (Contextualized Late Interaction over BERT) [8], [9], bilgi geri getirme sistemlerinde kelimelerin bağlamsal anlamlarına dayalı benzerlik bulma yöntemi sunmaktadır. Temel yenilik, geç etkileşim mekanizmasıdır. Diğer yöntemler, erken etkileşim stratejilerine dayanır; yani istem ve dokümanlar erken safhada birleştirilir ve bu, arama görevlerinde hesaplama karmaşıklığını artırabilir. ColBERT, her doküman

ve istemi ayrı işler ve etkileşim son aşamada gerçekleşir. Bu yapı, özellikle büyük veri kümelerinde daha verimli ve etkili sonuçlar sunan bellek dostu bir nitelik kazandırır. Diğer modellerin birçoğu, bellek kullanımı açısından daha maliyetlidir. TinyBERT gibi hafif modeller ise performans açısından zayıf kalmaktadır. Özellikle büyük ölçekli ve alan bazlı veri kümeleri üzerinde, anlamsal detayların korunması ve hassas arama sonuçları elde edilmesi için etkili olduğu görülmektedir.

Bu çalışmada üç farklı model, MSMARCO [10] veri kümesi kullanılarak eğitilmektedir. Veri kümesi, kullanıcıların sağlık ve tıbbi konular hakkında sordukları gerçek sorular ve açıklamalarını içermektedir. Bu sorular, belirli hastalıklar, semptomlar, tedavi yöntemleri, ilaçlar, tıbbi süreçler ve genel sağlık bilgileri gibi çeşitli konuları kapsamaktadır. İstem-bağlam ikilileri, bu veri kümesi kullanılarak oluşturulmaktadır. Modele eğitimde yanlış örnek olarak verilecek negatif örnekler, "multilingual-e5-large" modeli kullanılarak oluşturulmaktadır. Eğitimde, bilgi damıtma yöntemi kullanılarak veri kümesinde yüksek performanslı modeller elde edilmekte ve test aşamasında ColBERT yaklaşımı ile gömme vektörleri üretilmektedir. ColBERT'in geç etkileşim mekanizması sayesinde, üretilen vektörler arasındaki ilişki daha verimli bir şekilde analiz edilmekte ve bilgi geri getirme süreci hızlandırılmaktadır.

II. YÖNTEM

A. Genel Bakış

Bu çalışmada, alana özgü ince ayar yapılarak performansı ölçülen üç farklı modelin sonuçlarının değerlendirilmesi ele alınmaktadır. Çalışmanın ilk aşamasında, kullanılan veri kümesinin oluşturulması üzerinde durulmaktadır. Veri kümesinin hangi kriterlere göre üretildiği; ön işleme süreçleri, temizleme ve düzenleme işlemlerinden bahsedilmektedir. İkinci aşamada, oluşturulan veri kümesi kullanılarak modellere alana özgü ince ayar yapılması detaylandırılmaktadır. Üç modelin belli bir alana nasıl optimize edildiği ve performans artırımı için uygulanan teknikler yer almaktadır. Son olarak, veri kümesinden bilgi geri getirme işlemlerinde ColBERT yönteminin uygulanışı açıklanmaktadır. ColBERT yönteminin temelleri, bu çalışmadaki kullanımı ve sonuçların değerlendirilmesi ele alınmaktadır.

B. Veri Kümesinin Oluşturulması

Bu çalışmada, her biri insan tarafından oluşturulmuş istemlerden ve bağlamlardan oluşan MS-MARCO [10] veri kümesi kullanılmaktadır. Veri kümesi Türkçe diline çevrildikten sonra tıp alanına özgü veriler seçilmektedir. Bu seçim, tıp alanında derinlemesine bilgi sağlamak amacıyla yapılmaktadır.

Oluşturulan veri kümesi, satır geçişleri, yazım hataları ve noktalama işaretleri açısından düzenlenerek, gömme işlemi için optimize edilmektedir. Bu düzenlemeler, verilerin yapısını koruyarak modelin daha doğru ve güvenilir sonuçlar üretmesine katkı sağlamaktadır.

C. Seçilen Modellere Alana Özgü İnce Ayar Yapılması

Bu çalışma için üç tane model seçilmektedir.[1] Bu modeller, farklı dillerdeki metinleri işleme yetenekleriyle öne çıkmaktadır. Çalışmanın amacı doğrultusunda, bu modellerin performansını iyileştirmek için alana özgü ince ayar işlemi uygulanmaktadır.

Alana özgü ince ayar, modellerin tıp gibi özel bir alanda daha yüksek doğrulukla çalışmasını sağlamak amacıyla yapılmaktadır. Tıp alanında oluşturulmuş istemler ve ilgili bağlamlar çiftler olarak modele verilmektedir. Bu işlem, modellerin mevcut kapasitelerini artırarak belirli bir alan için özelleştirilmiş bilgi işleme yeteneklerini geliştirmektedir. Böylece, seçilen modellerin her biri, tıp alanında daha doğru ve güvenilir sonuçlar getirebilmektedir.

İnce ayar süreci, modellere istem ve istemle ilgili bağlam verilmesinin yanında zor negatif örnek madenciliği (Hard Negative Mining) tekniğini içermektedir. Bu teknikte, isteme göre hatalı olan ve modelin ayırt etmekte zorlandığı bağlamlar negatif örnek olarak belirlenmektedir. Zor negatif örneklerin tespiti için gömme modeli[2] kullanılmakta ve her istem için 10 yakın yanlış bağlam seçilmektedir. Bu sayede, <istem, pozitif-bağlam, negatif-bağlam> şeklinde bir eğitim veri kümesi oluşturularak, modelin daha karmaşık ve zorlayıcı örnekler üzerinde öğrenme kapasitesi artırılmaktadır.

D. Veri Kümesinden Geri Getirme

ColBERT (Contextualized Late Interaction over BERT) yöntemi, bilgi geri getirme işlemlerinde derin öğrenme modellerini kullanarak yüksek doğrulukta sonuçlar elde etmeyi amaçlayan bir yaklaşımdır. Bu yöntem, BERT (Bidirectional Encoder Representations from Transformers) modelinin dil anlayışını kullanarak metinlerin daha iyi analizinin yapılmasını sağlamaktadır. ColBERT, istem ve bağlam ilişkilerini verimli bir şekilde hesaplamak için geç etkileşim paradigmasını kullanmaktadır. Bu paradigma, istem ve bağlam temsillerini ayrı ayrı hesaplayıp sonrasında bu temsillerin etkileşimlerini mimarisindeki hızlı mekanizmayı kullanarak değerlendirir. faydalanmaktadır. Böylece büyük veri kümeleri üzerinde hızlı ve etkili bilgi geri getirme işlemi gerçekleştirmektedir.

Bu çalışmada, ColBERT'in metinsel analiz ve geç etkileşim paradigması, veri kümesinden bilgi getirme işlemlerinde kullanılan kosinüs benzerliği (cosine similarity) yöntemi ile desteklenmektedir. Kosinüs benzerliği, iki vektör arasındaki açıyı ölçerek benzerlik hesaplamakta ve özellikle metin tabanlı veri kümesi işlemlerinde kullanılmaktadır. Hesaplanan kosinüs benzerliğinin getirdiği sonuç, 1'e ne kadar yakınsa, vektörler o kadar benzer; -1'e ne kadar yakınsa, vektörler o kadar farklı kabul edilmektedir. Kosinüs benzerliğini ifade eden matematiksel fonksiyon, (1)'de verilmektedir.

$$\text{Kosinüs Benzerliği}(A, B) = \frac{\sum_{i=0}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

ColBERT mimarisindeki yenilik, her bir token için ayrı bir gömme vektörü üretmesidir. Bu çalışmada, geri getirme performansını ölçmek amacıyla, alana özgü veri kümesiyle ince ayar yapılmış gömme modelleri kullanılmaktadır. Veri kümesi ve istemlerin her bir token'ına ait gömme vektörleri hesaplanarak, istemlerden çıkarılan vektörlerle veri kümesinden çıkarılan vektörler arasında kosinüs benzerliği hesaplanmaktadır.

[1] Sırasıyla: "intfloat/multilingual-e5-base", "dbmdz/bert-base-turkish-cased" ve "sentence-transformers/distiluse-base-multilingual-cased-v1"

[2] "intfloat/multilingual-e5-large"

III. DENEYSEL ÇALIŞMALAR

Bu bölümde, bahsedilen üç modelin performans karşılaştırılması yapılmaktadır. Karşılaştırma, temel modeller ile alana özgü ince ayar yapılmış hallerinin başarımları üzerinden değerlendirilmektedir. İnce ayar sürecinde, modellere girdi olarak verilen veri kümeleri için toplu işlem boyutu (batch size) 8 olarak seçilmiştir. Ayrıca, modelin vektör sıkıştırması için kullanılan bit sayısı (nbits) 4 ve öğrenme oranı (learning rate) 0.000005 olarak ayarlanmıştır. Bu hiperparametreler, eğitim sürecini ve performansı etkileyen kritik faktörlerdir ve eğitilen tüm modeller için sabit tutulmaktadır.

İnce ayar yapılan modellerde zor negatif örnek madenciliği yapılırken iki farklı yöntem uygulanmaktadır. 10 zor negatif örnek üretilirken kullanılan modele sırasıyla iki farklı hibrit skor ağırlığı (weights) parametresi girilmektedir: 1 ve 0.5. Hibrit skor ağırlığı, modelin zor negatif örnekleri bulurken yaptığı aramanın derinliğini belirtmektedir. Değer 0'a eşitse sadece arama terimlerinin metinde bulunup bulunmadığını kontrol etmektedir. Değer 1 ise semantik arama yaparak daha detaylı bir analiz yapmaktadır. Örneğin, arama sonuçlarını kelime eşleşmesine göre değil, içerik benzerliğine ve anlamsal bütünlüğe bakarak kontrol etmektedir. Bu değeri 0.5 yaptığımızda ise hem basit kelime aramasını hem de hibrit arama kriterlerini kullanarak dengeli bir arama yapmaktadır.

Modellerin bilgi geri getirme performansı ölçülürken sırasıyla MRR (Ortalama Karşılıklı Sıra) ve İsabet Oranı (Hit Rate) metrikleri kullanılmaktadır. MRR, doğru sonucun sıralamada ilk bulunduğu yerin ters çevrilmiş değerlerinin ortalamasını almaktadır. Yüksek bir MRR skoru, modelin doğru yanıtları yüksek sıralarda getirdiğini göstermektedir. MRR skoru için kullanılan matematiksel formül, (2)'de verilmektedir. İsabet Oranı, belirli bir sayıdaki sonuçlar kümesinde doğru yanıtların bulunma sıklığını ölçmektedir. Belirli bir üst sınırdaki sonuçlarda doğru yanıt bulunma sıklığını hesaplamaktadır. İsabet Oranı skoru için kullanılan matematiksel formül, (3)'te verilmektedir.

$$\text{MRR} = \frac{1}{S} \cdot \sum_{i=1}^{S} \frac{1}{Derece_i} \tag{2}$$

MRR formülünde, S toplam istem sayısını, $Derece_i$ ise i. istemdeki doğru cevabın kaçıncı sırada olduğunu belirtmektedir.

$$\text{İsabet Oranı} = \frac{\textit{İsabet Sayısı}}{S} \tag{3}$$

İsabet Oranı formülünde $\textit{İsabet Sayısı}$ belirli bir sonuç kümesinde getirilen doğru yanıtların sayısını, S ise toplam istem sayısını ifade etmektedir.

Bu çalışmada seçilen üç modelin bilgi geri getirme performansı üç aşamada incelenmektedir. Bunun için istem ve bağlamlardan oluşan test veri kümesi tüm aşamalarda her modele ayrı ayrı verilmektedir. Ardından doğru cevabın sırasıyla ilk 1, 3, 5, 10, 20, 50 ve 100 tane belgenin içerisinde olup olmadığı kontrol edilmekte ve bunlara göre MRR ve Hit Rate skorlarına bakılmaktadır.

İlk olarak her modelin temel hallerinin performansları incelenmektedir. Burada üç modelde de herhangi bir şekilde ince ayar işlemi uygulanmadan test veri kümesi üzerinden skorlarına bakılmaktadır. Bu skorlar aşağıda Tablo I'de gösterilmektedir.

TOP K	multilingual-e5-base		dbmdz-turkish-bert		distiluse	
	MRR	Hit Rate	MRR	Hit Rate	MRR	Hit Rate
TOP 1	**0.437**	**0.437**	0.082	0.082	0.199	0.199
TOP 3	**0.519**	**0.623**	0.108	0.143	0.254	0.326
TOP 5	**0.537**	**0.699**	0.117	0.181	0.266	0.381
TOP 10	**0.547**	**0.770**	0.124	0.234	0.277	0.460
TOP 20	**0.551**	**0.828**	0.129	0.305	0.282	0.539
TOP 50	**0.552**	**0.872**	0.132	0.393	0.285	0.638
TOP 100	**0.553**	**0.892**	0.133	0.455	0.286	0.699

Tablo I: Temel Modellerin Skorları

İkinci aşamada aynı üç modele ince ayar yapılmaktadır. Bu süreçte modellere, istemler ve bu istemlerle ilgili bağlamlardan oluşan bir eğitim veri kümesi verilmektedir. Eğitim ve test veri kümelerinin farklı olmasına dikkat edilmektedir. Ayrıca ince ayar sırasında tüm modeller için zor negatif örnekler üretilirken hibrit skor ağırlığı parametresi 1 olarak ayarlanmaktadır. İnce ayar sonrasında modellerin MRR ve Hit Rate skorlarında artışlar gözlemlenmektedir. Özellikle ince ayar yapılmış *"dbmdz-turkish-bert"* modelinde belirgin bir performans artışı kaydedilmektedir. Bu performans sonuçları, aşağıda Tablo II'de yer almaktadır.

TOP K	multilingual-e5-base		dbmdz-turkish-bert		distiluse	
	MRR	Hit Rate	MRR	Hit Rate	MRR	Hit Rate
TOP 1	0.444	0.444	**0.509**	**0.509**	0.250	0.250
TOP 3	0.534	0.650	**0.598**	**0.708**	0.309	0.386
TOP 5	0.551	0.724	**0.615**	**0.785**	0.325	0.455
TOP 10	0.562	0.780	**0.625**	**0.854**	0.337	0.547
TOP 20	0.566	0.857	**0.628**	**0.903**	0.343	0.633
TOP 50	0.571	0.887	**0.629**	**0.937**	0.346	0.715
TOP 100	0.569	0.889	**0.630**	**0.953**	0.347	0.760

Tablo II: İnce Ayar İşlemi Yapılmış Modellerin Skorları

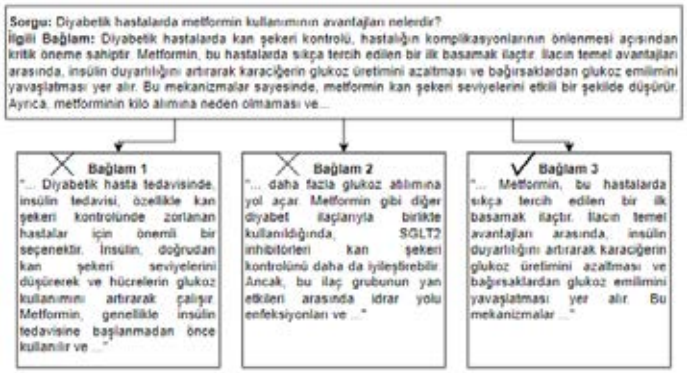
Son aşamada, ikinci aşamaya benzer bir ince ayar işlemi yapılmaktadır. Ancak ikinci aşamadan farklı olarak zor negatif örnekler oluşturulurken hibrit skor ağırlığı 0.5 olarak verilmektedir. Eğitilen modeller aynı test veri kümesinde değerlendirilmektedir. Sonuçlara göre, *"multilingual-e5-large"* modelinin performansında ikinci aşamaya göre artış gözlemlenmektedir. Diğer iki modelde ise çok az bir skor değişikliği kaydedilmektedir. Bu skorlar Tablo III'te listelenmektedir.

Elde edilen bulgulara göre MRR ve Hit Rate skorları, ince ayar işleminin modellerin doğru sonucu bulma ve doğru sonucu ilk sıralarda getirme yeteneğini artırdığını göstermektedir. Ayrıca zor negatif örneklerin oluşturulmasında kullanılan hibrit skor ağırlığı parametresinin, özellikle *"multilingual-e5-large"* modelinde önemli bir performans artışı sağladığı sonucuna ulaşılmıştır.

Şekil 1 ve Şekil 2'deki örnekte bir istem, istemle ilgili

TOP K	multilingual-e5-base		dbmdz-turkish-bert		distiluse	
	MRR	Hit Rate	MRR	Hit Rate	MRR	Hit Rate
TOP 1	**0.526**	**0.526**	**0.526**	**0.526**	0.250	0.250
TOP 3	0.607	0.708	**0.611**	**0.720**	0.308	0.386
TOP 5	0.624	0.783	**0.627**	**0.788**	0.321	0.441
TOP 10	0.633	0.849	**0.637**	**0.856**	0.333	0.532
TOP 20	0.636	0.895	**0.641**	**0.912**	0.338	0.608
TOP 50	0.637	0.937	**0.642**	**0.947**	0.341	0.688
TOP 100	0.638	0.951	**0.642**	**0.960**	0.342	0.747

Tablo III: İnce Ayar İşlemi Yapılmış Modellerin Skorları



Şekil 1: Temel Modelin Bilgi Getirme Örneği

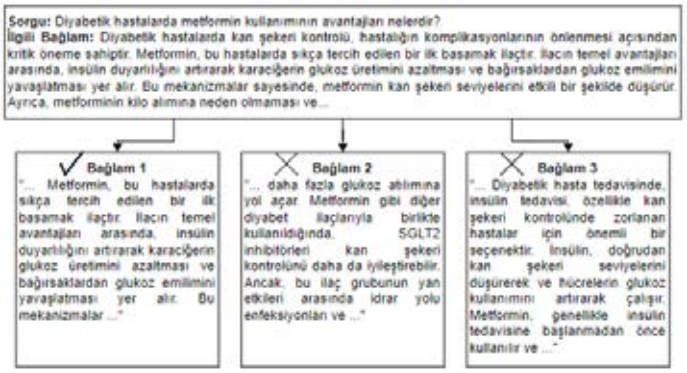
bağlam ve bu isteme göre model tarafından getirilmiş üç tane belge gösterilmektedir. Şekil 1'deki örneğe göre temel model kullanılırken doğru cevap üçüncü sıradaki belgede getirilmektedir. Ardından Şekil 2'deki örnekte ince ayar işlemi uygulanmış model kullanılmakta ve doğru cevap birinci sırada getirilmektedir. Bu da ince ayar işleminin modelin bilgi getirme performansını artırdığını göstermektedir.

IV. SONUÇLAR VE TARTIŞMA

Bu çalışmada, bilgi geri getirme işlemlerinde tıp alanına özgü ince ayar süreçlerinin model performansına etkisi detaylı bir şekilde incelenmektedir. Temel modellerin performansı ile zor negatif örnek madenciliği kullanılarak ince ayar yapılan modellerin performansı MRR ve Hit Rate skorlarına bakılarak karşılaştırılmaktadır. Elde edilen sonuçlar, ince ayar işleminin modellerin performansında artış sağladığını göstermektedir. Bu bulgular, ince ayar sürecinin modellerin tıp gibi uzmanlık gerektiren alanlarda daha etkin bilgi geri getirme yeteneği kazanmalarına olanak sağladığını göstermektedir.

Çalışmanın sonuçları, tıp alanındaki bilgi geri getirme sistemlerinin performansını artırma potansiyelini ortaya koymaktadır. İnce ayar işlemi sonrasında elde edilen yüksek skorlar, klinik veri tabanları ve hastane bilgi yönetim sistemleri gibi alanlarda daha isabetli ve hızlı bilgi erişimi sağlayabileceğini göstermektedir.

Bununla birlikte, çalışmanın bazı sınırlamaları bulunmaktadır. İlk olarak, kullanılan eğitim ve test veri kümeleri tıp alanına özgü olarak hazırlanmış olsa da, veri kümesinin boyutu ve çeşitliliği sınırlı kalmaktadır. Daha geniş ve çeşitli veri



Şekil 2: İnce Ayar Yapılmış Modelin Bilgi Getirme Örneği

kümeleriyle gerçekleştirilecek deneyler, modellerin daha iyi performans sunup sunamayacağını ortaya koyabilir. İkinci olarak, hibrit skor ağırlığı parametresinin belirlenmesinde yalnızca iki farklı değer test edilmektedir. Gelecekteki çalışmalar, bu parametrenin farklı ayarlarını da test ederek en uygun değeri belirleyebilir.

Bu çalışma, zor negatif örnek madenciliği ve hibrit skor ağırlığı parametresi gibi tekniklerin bilgi geri getirme performansını artırmada ne denli etkili olabileceğini göstermesi açısından literatüre katkı sunmaktadır. Gelecekte yapılacak çalışmalar, bu yöntemlerin farklı veri türleri ve daha büyük veri setleri üzerinde uygulanmasını inceleyerek, elde edilen sonuçları geliştirebilir.

KAYNAKÇA

[1] Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Urvashi Khandelwal, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, and Sebastian Riedel. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." arXiv preprint arXiv:2005.11401 (2020).

[2] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. "Retrieval-Augmented Generation for Large Language Models: A Survey." arXiv preprint arXiv:2312.10997 (2023).

[3] Devlin, Jacob. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[4] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations, 2020.

[5] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[6] Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. Tinybert: Distilling bert for natural language understanding, 2020.

[7] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks, 2019.

[8] Omar Khattab and Matei Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over bert, 2020.

[9] Keshav Santhanam, Omar Khattab, Jon Saad-Falcon, Christopher Potts, and Matei Zaharia. Colbertv2: Effective and efficient retrieval via lightweight late interaction, 2022.

[10] Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. Ms marco: A human generated machine reading comprehension dataset, 2018.