

Dokumentumfeldolgozáson alapuló magyar nyelvű chat AI RAG technológiával

Önálló laboratórium 2 – Szladek Máté (TGPZTT)

Konzulensek: Gyires-Tóth Bálint és Ónozó Livia

<https://github.com/TGPZTT/Onlab2>

2024.05.21.

FELADAT BEMUTATÁSA

- Nem feltétlenül strukturált kérdések
- Pontos válaszok
- Hivatkozások (forrás)
- Magyar nyelvű válasz (lehet angol a forrás)

Probing for the Trace Estimation of a Permuted Matrix Inverse
Corresponding to a Lattice Displacement*

Heather Switzer[†] Andreas Stathopoulos[‡] Eloy Romero[§]
Jesse Laeuchli[¶] Kostas Orginos^{||}
November

Probing [30] is a general technique for estimating the trace of a matrix A^{-1} [22]. The variance of the estimator is used in the estimator they annihilate the elements of A^{-1} . Therefore, this method is used to display decay of the magnitude of the elements of A^{-1} . Equivalently, when a matrix is used to find a distance, coloring was proposed so that the variance reduction over conventional probing methods. The rest of the paper is organized as follows: Section 2 discusses previous variance reduction techniques. Section 3 discusses the method with displacements, and studies its properties theoretically. Section 4. Conclusions and some open questions.

2 Background

In this paper we seek the trace of PA^{-1} , where P is a permutation matrix and A is a non-singular matrix of dimension N . A can be used for convenience and without loss of generality. Although our main idea applies to any matrix, we focus on the case where A is a Laplacian matrix, which is relevant to matrices stemming from the multiplicative group of integers modulo N .

1 Introduction

The approximation of A^{-1} is a computational problem. The approximation of $\log A^{-1}$ is a data mining [8]. Our method is based on the LQCD.

*This work is supported by the National Science Foundation (NSF) Grant PHY-17-09891 and the National Science Foundation (NSF) Grant PHY-17-09891. †Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ‡Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. §Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ¶Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ||Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

where D_i is the size of dimension i and x_i are their coordinate vectors $[x_1, \dots, x_n]$ (in the sense). In LQCD, the lattice representation of the system is used.

Variance reduction techniques are used to address the problem of the approximation of A^{-1} . One of the better methods is the Hadamard or Fourier matrix method, which is based on the specific diagonals of the matrix. The method is based on the use of random noise vectors as the pattern of the contributing diagonals of the matrix.

2.1 Classical

The inverse of an $n \times n$ matrix A can be computed using the Neumann series $A^{-1} = \sum_{k=0}^{\infty} (I - A)^k$, where I is the identity matrix. This method is used to compute the inverse of a matrix.

Partial Differential Equations, Lattice QCD, and the LQCD method are used to compute the inverse of a matrix. The LQCD method is used to compute the inverse of a matrix.

The CP (classical probing) method is not used to compute the inverse of a matrix. The CP method is used to compute the inverse of a matrix.

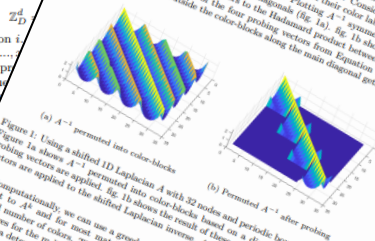


Figure 1: (a) A^{-1} permuted into color-blocks. (b) Permuted A^{-1} after probing. The figure shows the trace of A^{-1} as a function of the distance k . The trace is approximately $1/k$. The plots show the trace of A^{-1} as a function of the distance k . The trace is approximately $1/k$.



MEGOLDÁSOK

Kulcsszavas keresés

Nagyon egyszerű
Könnyen implementálható
Kontextus nem vizsgál, pontatlan

Szemantikus keresés

A szöveg jelentését is ismeri
(vektoros reprezentációk)
Módszertani korlátok: Lexikai homonímia
Out-of-vocabulary

Q&A rendszerek

Információkinyerési technológia
Mély tanulás megjelenésével
óriási fejlődésnek indult

Témamodellezés

Transzformer modellek
fine-tuningolása

Még sok más...

RAG - Retrieval Augmented Generation

- Generatív- és információ-visszakeresési (IR) modell ötvözése
- Egyszerre ad választ és tanul is. (A legtöbb megoldás külön lépésekben csinálja)
- Kontextusérzékeny
- Iteratív válaszfinomítás
- Tudásintegrálás – a nyelvi modell meglévő tudása + a visszakeresett információk

FELHASZNÁLT KOMPONENSEK

- arxiv
- FIASS vektortároló (META – facebook)
- RAG-os LangChain
- Gradio chatbot

```
## Start kérdés
test_question = "Tell me about Hungary!"
```



```
{'input': 'Tell me about Hungary!',
 'history': '', 'context': '[Quote
from Digital Economy And Society. A
Cross Country Comparison Of Hungary
And Ukraine] Table 2. Connectivity
scores Source: EC (2017)\nHungary
ranks 14th in this component,
compared to\n16th in 2016. Hungary
has made progress both in the
supply \nand the demand side. Fast
broadband coverage increased to
\n81% from 78%. The Hungarian
government launched two
\ninitiatives to increase demand
[redacted]
} }
```

Sure, I'd be happy to tell you about Hungary! When it comes to the digital economy and society, Hungary has made progress in recent years. According to the Digital Economy and Society Index (DESI), Hungary ranks 14th in connectivity, which is an improvement from 16th place in 2016. The country has seen an increase in fast broadband coverage, which now stands at 81%, and the government has launched initiatives to increase demand for broadband, such as a preferential VAT rate for subscriptions and a "digital welfare basic tariff" for non-users.

Hungary's human capital in terms of digital skills is a bit of a mixed picture. While only half of the population has at least basic digital skills, the country has a higher share of ICT specialists in the labor force (4.9%) compared to the EU average (3.7%). However, the study notes that integration of digital technology by businesses is a major issue in Hungary, and that Hungarian firms should make better use of online business, social media, and cloud-based applications.

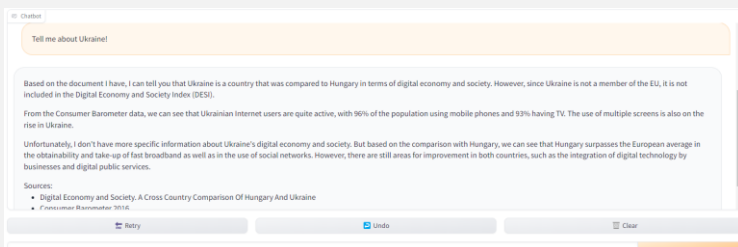
When it comes to e-commerce, the most popular device in Hungary is still the computer, but the popularity of smartphones and tablets is on the rise. The use of multiple screens is also becoming more common in Hungary.

As for the use of the internet, Hungary performs well above the European average. In fact, Hungarian internet users are avid consumers of online content, with 96% of the population using mobile phones and 93% having TV. The ratio of STEM graduates per 1000 individuals aged 20-29 is quite low in Hungary, but the government has implemented a new Digital Competences Strategy to increase the ratio of STEM graduates and address lifelong learning.

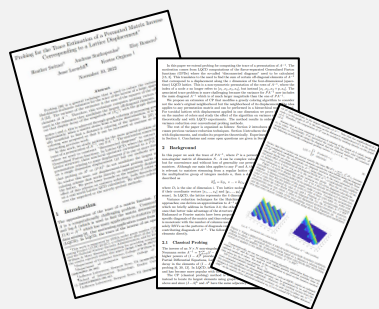
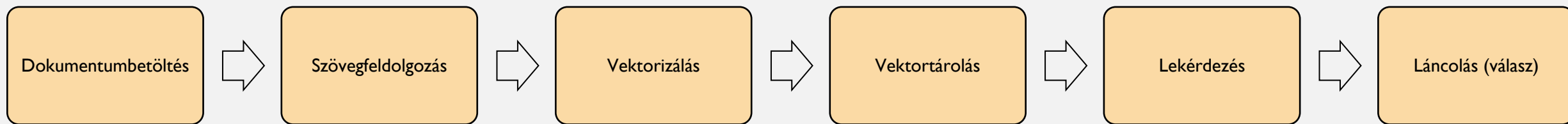
Overall, Hungary's DESI score is above the EU average in the use of internet and somewhat below the average in connectivity and human capital. The country needs to improve the integration of digital technology by businesses and digital public services.

Sources:

- * Digital Economy And Society. A Cross Country Comparison Of Hungary And Ukraine
- * E-commerce in Hungary: A Market Analysis
- * The Digital Scoreboard (EC 2016b)
- * Consumer Barometer 2016



RAG FOLYAMAT



Akár különböző
forrásokból

RecursiveCharacterTextSplitter

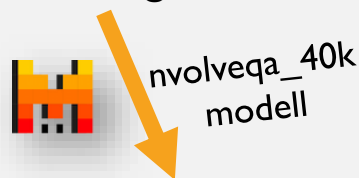
Szöveg da rabolás



1000 karakter
méretű, de átfedő
szövegszeletek

Ha $X < 200$ akkor kuka

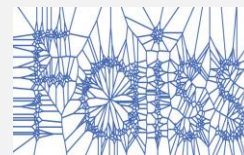
Szöveg



numerikus vektorok

Vektorok:

- Szövegrészek
- Metaadatok
- Extra információk (cím)



- A kérdésnek megfelelő
max 5000 karakteres
válaszcsoomag a
vektortárolólóól

```
{'input': 'Tell me  
about Hungary! ',  
'history': '',  
'context': '[Quote  
from Digital Economy  
And Society. A Cross  
Country Comparison Of  
Hungary And Ukraine]  
Table 2. Connectivity  
scores Source: EC  
(2017)\nHungary ranks  
14th in this  
component, compared  
to\n16th in 2016.  
Hungary has made  
progress both in the  
supply '}]
```

- Prompt engineering
- **Input:** kérdés +
válaszcsoomag
- **Output:**
Megfogalmazott válasz
- Ha nincs válasz a dokukban,
akkor a modell önálló tudását is
használhatja, csak akkor nem
hivatkozik

DOKUMENTUM - CHATBOT



embender



- ai-mixtral-8x7b-instruct
- nvolveqa_40k
- LangChain

Képes nagy bemeneteket kezelni (llama3 nem)

Stream chain: chat prompt | llm

Angol válasz



mixtral_8x7b



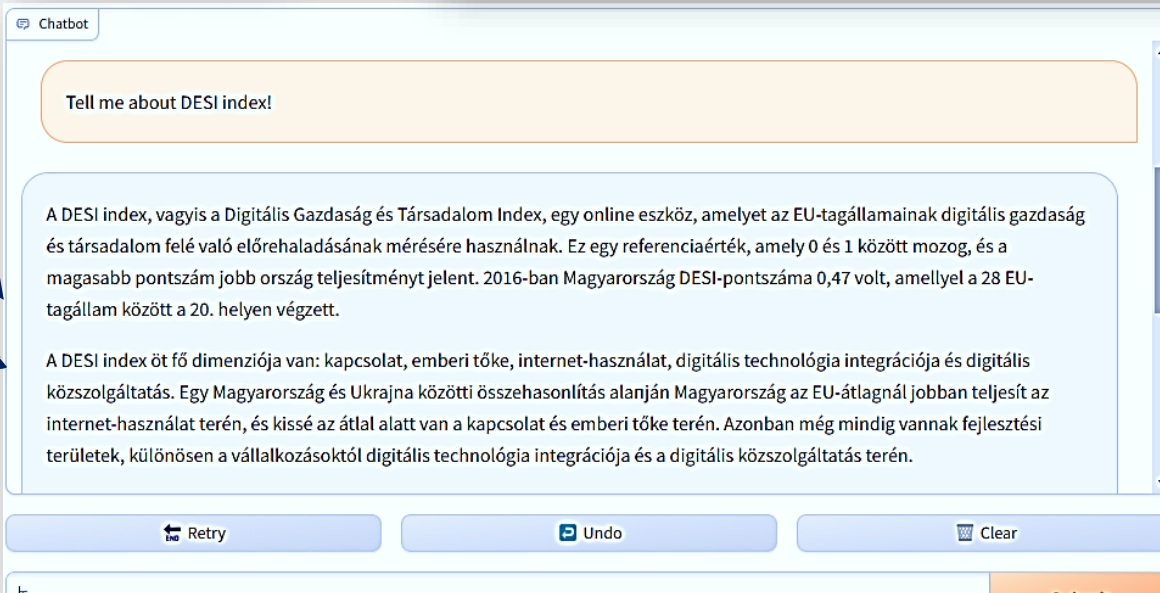
LE CHAT MISTRAL_

Fordító modellnek: llama3

Promptolással jó nyelvezetet ad.

```
print(translate("This apple is red, but not green. "))
```

Ez az alma piros, de nem zöld.



API key
kérdés

KÖSZÖNÖM A FIGYELMET!