

## Matematika- és informatikatudomány

		<b>Informatikatudomány I.</b>	<b>Matematikatudomány</b>
		<b>Szekcióelnök: Bodroginé Dr. Zichar Marianna (DE)</b>	<b>Dr. Bessenyei Mihály (DE)</b>
<b>I. panel</b>	<b>10.30-10.45</b>	Al-Kuran Hussam	Beke Ákos
	<b>10.45-11.00</b>	Bognár Eszter Katalin	Borsos Benjámín
	<b>11.00-11.15</b>	Gera Imre	Janabi, Hayder Abbas
	<b>11.15-11.30</b>	Ghanim Hussein Ali Ahmed	Kovács Bence Máté
	<b>11.30-11.45</b>	Sewunetie Walelign Tewabe	
		<b>Informatikatudomány II.</b>	
		<b>Szekcióelnök: Dr. Herendi Tamás (DE)</b>	
<b>II. panel</b>	<b>13.00-13.15</b>	Hajdu Flóra	
	<b>13.15-13.30</b>	Gavua Ebenezer Komla	
	<b>13.30-13.45</b>	Mattyasovszky-Philipp Dóra Anna	
	<b>13.45-14.00</b>	Padányi Viktória	
	<b>14.00-14.15</b>	Ősz Olivér	
	<b>14.15-14.30</b>	Barth Áron	

Informatikatudomány I.

2020. október 16. 10.30-13.15

Szekciófelelős:

Forczek Krisztofer, +36 30 388 0787

### **3D image preprocessing through morphological transformation**

**AL-KURAN**, Husam

Szent Istvan University

Computer Vision

*Husam.al-kuran@phd.uni-szie.hu*

Modern 3D imaging devices can potentially revolutionize the automated farm produce sorting industry. 3D imagers provide more insight into fruit shape and size in addition to conventional visible light information. However, 3D imagers produce large amounts of data for each image. Such data need to be pre-processed to reduce its size and to eliminate dependence on posture without sacrificing quality.

In this paper, a method for 3D image transformation is introduced than can improve compatibility with conventional Artificial Neural Networks (ANNs) through mapping the surface of a fruit onto a map of a sphere akin to topological maps used in geology. The method also allows conventional image compression technologies to be applied to 3D images.

2020. október 16. 10.30-13.15

Szekciófelelős:

Forczek Krisztofer, +36 30 388 0787

### **Migrációs trendek feltérképezése adatbányászat segítségével**

**BOGNÁR** Eszter Katalin

NKE KMDI

Védelmi elektronika, informatika és kommunikáció

*bognarek@uni-nke.hu*

Napjaink egyik legégetőbb problémája a migrációs válság, más néven európai menekültválság. A kifejezést először 2015-ben használták, mikor a különböző fegyveres konfliktusok, a politikai és vallási üldöztetés, gazdasági ellehetetlenülés miatt nagyszámú menekült indult útnak főként a Közel-Keletről, Afrikából, a Balkánról és Közép-Ázsiából, hogy a jobb élettel kecsegtető Európai Unió területére jusson. A főbb migrációs útvonalak a Földközi-tengeren és a Balkánon alakultak ki, és a balkáni útvonal révén hazánk is nagy mértékben érintett lett a problémában. A migráció háttere rendkívül sokrétű, a háborútól és üldözéstől való félelem mellett ugyanolyan hajtóerőként szerepel a létbizonytalanságból való kitörés szándéka, illetve a „nyugati” életszínvonal elérésének reménye. Ennek a komplex folyamatnak az átláthatóvá tétele érdekében elengedhetetlen a publikusan is elérhető adatbázisok és statisztikák tanulmányozása, az információbőség ugyanakkor rendkívül nehézkessé teszi a főbb trendek azonosítását, a lényeg megértését.

A problémára megoldást nyújthat az adatbányászat, melyet már számos területen sikeresen alkalmaznak, elsősorban az üzleti és pénzügyi elemzések, marketing és értékesítés, diagnosztika, gyártási folyamatok területén. Egy újabb terület, ahol az adatbányászatban rejlő lehetőségeket szintén ki lehet aknázni, a társadalmi problémák azonosítása és folyamatok feltérképezése, ez által hatékonyabb társadalmi döntések meghozatalának lehetővé tétele. Előadásom célja ismertetni az adatbányászat alkalmazásának lehetőségét a migrációs trendek feltérképezése céljából. Az előadás bemutat egy lehetséges elemzési folyamatot, beleértve az elemzés alapjául szolgáló adatforrások kiválasztását és előfeldolgozását, a szükséges technológiai háttérrel, valamint ismerteti az elemzés során nyert legjelentősebb eredményeket.

2020. október 16. 10.30-13.15

Szekciófelelős:

Forczek Krisztofer, +36 30 388 0787

### **Korrelációmátrix-szűrési eljárások a Markowitz-portfóliómodell javítására**

**GERA Imre, LONDON András**

Szegedi Tudományegyetem

Informatika, Közgazdaságtudomány

*gerai@inf.u-szeged.hu, london@inf.u-szeged.hu*

Az elmúlt években több tanulmány is részletesen elemezte a részvénytársasági hozamok idősorából képzett kovariancia- és korrelációmátrixokat és azok úgynevezett szűrési eljárásokkal történő módosítását. A szűrés fő motivációja, hogy a kovarianciamátrixokat csak becsülni tudjuk, és mivel az idősoraink is csak végesek, emiatt a becslések gyakran hoznak magukkal statisztikai bizonytalanságot vagy másnéven zajt. Ebben a munkában több megközelítést is bemutatunk, amelyek a korrelációmátrixokban előforduló zaj mértékét próbálják meg csökkenteni: a két fő irányvonal a véletlen mátrixok elmélete, illetve a hierarchikus klaszterezés.

A véletlen mátrixok elméletén alapuló szűrésnél egy, az eredeti korrelációs mátrixhoz hasonló tulajdonságú, ám véletlen korrelációs mátrixot generálunk, majd eredeti (becsült) mátrix és a véletlen mátrix sajátértékeit hasonlítjuk össze. A szűrés megvalósítására ezután főkomponens-elemzést használhatunk.

A hierarchikus klaszterezés esetében a korrelációs mátrixokra, mint gráfok szomszédsági mátrixaiként (részvénygráf) tekintünk, így a hálózattudomány eszközeit is felhasználhatjuk. Ilyenkor a részvénygráf egy tömörebb reprezentációját keressük, például egy minimális feszítőfáját. Az eljárás lényege, hogy a korrelációs mátrixot először egy távolságmátrixszá alakítjuk (egy ultrametrikus távolság felhasználásával), ahol a kis távolságok nagy korrelációra utalnak. Ezt a távolságmátrixot gráfként értelmezve minimális feszítőfát kereshetünk, így csak a legfontosabb korrelációk és a gráf összefüggősége marad meg. Ez a gráf végül visszaírható egy új, „szűrt” korrelációs mátrixszá.

Természetes lenne továbbá ezeken a részvénygráfokon más hálózatelemzési módszereket, például közösségkeresést alkalmazni, azonban több tanulmányban is rámutattak, hogy ezt közvetlenül a részvénygráfon alkalmazva torzított eredményeket kaphatunk. Mi ezt egy nullmodell alapú eljárással kerüljük el. A nullmodell szerint a páronkénti korrelációkat véletlennek tekintjük, azonban – amennyire lehetséges – nem változtatjuk a korrelációs mátrixban a sorösszegeket. Az eredeti és nullmodell mátrix közti különbség az elem párok közti korreláció „szignifikanciáját” méri. Ezután hierarchikus klaszterezést alkalmazunk a különbségmátrixra.

A szűrési eljárások teljesítményét a Markowitz-féle portfólióoptimalizálási feladaton mérjük le a Budapesti Értéktőzsde historikus adatainak felhasználásával. A hatékonyság megállapításához a szűrt kovarianciamátrixok segítségével és az eredeti modellel összeállított portfóliókon meghatározzuk a realizált hozamokat, a becsült és realizált kockázat hányadosát, a portfóliók Sharpe-hányadosát, valamint a módszerek által kiválasztott részvények számát. Egy további heurisztika portfólió összeállításra a klaszterező eljárások esetén, hogy a klaszterek számát előre megadva minden klaszterből csak 1-1 részvényt engedünk a portfólióba válogatni, így elkerülve az egymással erősen korreláló részvényeket.

Kísérleteink azt mutatják, hogy a Markowitz-modell minden szűrés esetében lényegesen kevésbé becsüli rosszul (alul) a kockázatot, illetve módszertől függően egy vagy több másik mutatója is javítható az összeállított portfólióknak az eredeti modellhez képest.

2020. október 16. 10.30-13.15

Szekciófelelős:

Forczek Krisztofer, +36 30 388 0787

## **Development of Recommendation Systems Based on Collaborative Filtering Using KNN and SVD**

**GHANIM HUSSEIN** Ali Ahmed, László **KOVÁCS**

Institute of Information Technology, University of Miskolc, Miskolc, Hungary

Ontology Model for Etutoring Systems

*ghanim@iit.uni-miskolc.hu, kovacs@iit.uni-miskolc.hu*

Educational Recommender System helps learners to identify the most interesting and relevant books from a large number of books. Data mining methods can be used to determine the relevant and useful knowledge sources from an existing document pool. This work focuses on relevance analysis of Data Mining methods with special reference to collaborative filtering in Educational Recommender Systems. Various tools are available which supports the related classification process like Recommendation system (RS). RS is a system which performs prediction for recommending some items to users; it filters the data and recommends some items. RS can be classified into three different categories based on the technique used; we can use content-based method, collaborative Filtering method or hybrid method. In this work we focus on Collaborative Filtering (CF) approach. CF is a method of identifying the similar users and recommending what the common users prefer. CF is based on the principle that similar users will share similar interests. CF algorithms can be divided into two categories; they are Memory-Based and Model-Based approaches.

Recommendation systems have well-known success in several domains, as in e-commerce and books recommendation. Some of the most successful recommendation techniques are K-Nearest Neighbors (KNN) and Singular Value Decomposition (SVD) collaborative filtering. KNN is classification algorithm, one of the most popular used non-parametric classification methods, however it is limited due to memory consumption related to the size of the dataset, which makes them impractical to apply to large volumes of data. SVD is one of the Matrix Factorization models for identifying hidden factors. We use combination of KNN and SVD to get better results. In this work we use item-based CF by using KNN and SVD algorithm. This work focuses on comparing two different collaborative filtering algorithms, in which the aim was to find out which one that produces the best prediction rate. The two investigate algorithms were KNN and SVD. According to our test result out of the two algorithms the SVD had the best prediction rate. Our results show that the approach to predict the books ratings can be used as a prediction engine for the recommendation system. In the future we can use this approach to develop a book recommendation system.

*Keywords: Recommendation Systems, Data Mining, Collaborative filtering, KNN, SVD.*

2020. október 16. 10.30-13.15

Szekciófelelős:

Forczek Krisztofer, +36 30 388 0787

### **Evaluation of Template Based Automatic Question Generation Techniques**

**SEWUNETIE WALELIGN** Tewabe, László **KOVÁCS**

University of Miskolc, József Hatvany Doctoral School of Computer Science and Engineering,  
Miskolc, Hungary

sewunetie@iit.uni-miskolc.hu, kovacs@iit.uni-miskolc.hu

A question is a sentence or clause in which the subject to be addressed by the student appears and is represented as part of the text as a whole. A question contains all the information related to the generation process related to a particular item stem, base, distractors, and data. A good question can ignite creativity and also educate students. While developing good questions may seem like a pretty straightforward task, it is not as simple as it looks. A good question needs to have a personality of its own which is made of specific characteristics. In the education field, the academicians are majorly dependent on their own for generating questions for various examinations. In the ongoing advancement of computer-based technology, there are many changes being changed that tend to move from manual systems to automated systems. The automatic systems help us with many costs and time-efficient solutions. Education is one of these areas which have gone through this evolution.

Question generation can help a person to generate questions from the given text automatically. The potential benefits of using automated systems to generate questions help reduce the dependency on humans to generate questions and other needs associated with systems interacting with natural languages. We have used Python's programming language and spaCy library to develop sample template based AQG application and we describe architecture, steps, and preliminary evaluation results. Even if there are different question generation approaches, in this study, we analyze and doing sample application using template based techniques of AQG. We understand that to ensure all questions that can be entered into the system and return a corresponding answer template based question generation is the best technique. The evaluation showed that system-generated questions could not be distinguished from human-generated questions in the context of two discussion topics while the difference between system-generated and human-generated questions was noticed in the context of one discussion topic. In addition, the study showed that templated based system-generated questions are limited in type and it is difficult to generate case based questions.

**Keywords:** Automatic Question Generation, Template Based, spaCy library, System Generated Questions.