

Aufgabenstellung

Definitive Anmeldung zum Wirtschaftsprojekt/zur Bachelorarbeit

1. Starttermin:

spätmöglicher Starttermin: HS KW 38; FS KW 8
17.02.2020

2. Abgabetermin:

Dauer einer BAA: max. 15/16 Kalenderwochen (Um zur ordentlichen Diplomierung im Sommer zugelassen zu werden, muss die Abgabe bis spätestens Freitag, eine Woche nach Semesterende, erfolgen)

Dauer eines WIPRO: max. 14/15 Kalenderwochen
05.06.2020

3. Studierende:

	Student/in 1:	Student/in 2:
Name, Vorname:	<u>Gröger Fabian</u>	
Studiengang:	<u>Informatik</u>	
Mobile:	<u>+41 79 554 93 01</u>	
E-Mail:	<u>Fabian.groeger@stud.hslu.ch</u>	
Projekt mit Arbeitgeber (bb-Studierende)	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein	<input type="checkbox"/> Ja <input type="checkbox"/> Nein

4. Auftraggeber/in:

Firma:	<u>HSLU</u>
Ansprechperson:	<u>Daniel Pfäffli</u>
Funktion:	<u></u>
Strasse:	<u></u>
PLZ / Ort:	<u></u>
Telefon:	<u></u>
Email:	<u></u>
Website:	<u></u>

5. Betreuende/r Dozent/in:

Daniel Pfäffli

6. Aufgabenstellung

Titel:	Deep embedded Music
Ausgangslage und Problemstellung:	<p>Die Unterscheidung von Musikstücken ist für den Menschen von einfach bis schwierig. So ist es meistens einfach Musikstilrichtung, wie Klassik oder Techno zu unterscheiden. Jedoch erhöht sich die Schwierigkeit, wenn wir innerhalb eines Styles Lieder vergleichen wollen oder diese sogar nach Ähnlichkeit sortieren wollen.</p> <p>Auf einem etablierten Datenset für Geräusch-Detektion soll ein Verfahren basierend auf einem Embedding-Verfahren mit Namen Tile2Vec erarbeitet werden, welches Similaritäten zwischen Geräuschen vergleichen kann. Auf den resultierenden Embeddings soll ein einfacher Klassifizierer trainiert und die Genauigkeit des Verfahrens den publizierten Resultaten gegenübergestellt werden.</p> <p>Dieses Verfahren soll explorative auf Musik unterschiedlicher Art angewandt werden, wobei ein Datensatz von 12h Jodelmusik zur Verfügung steht. Die entstehenden Cluster sollen qualitativ untersucht werden.</p>
Ziel der Arbeit und erwartete Resultate:	<p>Ziel der Arbeit ist es, Tile2Vec, ein Bild-Embeddingsverfahren, auf Audio Streams zu adaptieren und seine Performanz auf einem Geräusche-Detektions-Datenset «SINS, DCASE 2018 task 5 development dataset» zu evaluieren. Dabei soll mittels unsupervised Machine Learning ähnliche Signale «embedded» und darauf ein einfacher Klassifizierer trainiert werden. Das erarbeitete Embedding-Verfahren soll explorativ auf Musik angewandt und die resultierenden Cluster qualitativ untersucht werden.</p> <p>Gegeben:</p> <ul style="list-style-type: none"> • Datenkorpus «SINS, DCASE 2018 task 5 development dataset» • Embedding Verfahren Tile2Vec <p>Erwartetes Resultat:</p> <ul style="list-style-type: none"> • Ausführliches Study Doc • Schlussbericht gemäss Vorgaben des Bachelor-Studiengangs Informatik • Schlusspräsentation nach Abgabe der Arbeit (20 Minuten) • Evaluationsworkflow • Implementation von Tile2Vec in Tensorflow 2.0 • Empfehlung und Begründung eines geeigneten Preprocessing des quantisierten Signals. • Empfehlung und Begründung für eine Embedding-Architektur auf Basis von Neuronalen Netzwerken. • Identifikation eines geeigneten Similaritätsmass, um Nachbarschaften von «embedded» Audiosignalen finden zu können. • Qualitatives Resultat auf Songkorpus <p>Optionale Ziele;</p> <ul style="list-style-type: none"> • Evaluation mit unterschiedlichen Audiosignallängen für Embeddings auf dem SINS Datenset und aufzeigen des Einflusses. • Empfehlung und Begründung für geeignete Audiolänge und Preprocessing für Musik. <p>Abgrenzungen:</p> <ul style="list-style-type: none"> • Der Fokus der Arbeit liegt ausschliesslich auf der Architektur des Embedding-Modelles, sowie dem Preprocessing des Audiosignales.

	<p>Clustering und Klassifizierung wird explizit vom Stand der Forschung ausgeschlossen. Es soll stattdessen auf bewährte Standardverfahren zurückgegriffen werden.</p> <ul style="list-style-type: none"> • Optimierungsverfahren zum Trainieren von neuronalen Netzwerken wird explizit vom Stand der Forschung ausgeschlossen. Es soll stattdessen auf bewährte, in Tensorflow implementierte, Verfahren zurückgegriffen werden.
Gewünschte Methoden, Vorgehen:	<p>Vorgehen:</p> <ul style="list-style-type: none"> • Planung, Organisation, Risiko- und Ressourcenmanagement sind Teil der Aufgabe und werden von den Studierenden wahrgenommen • Während der Arbeit ist ein persönliches Arbeitsjournal zu führen. Ein Arbeitsjournaleintrag umfasst Datum, Anzahl Stunden und Arbeitsschritt/Thema. <p>Für folgende Teile muss eine Abnahme durch den Betreuer in die Planung aufgenommen werden:</p> <ul style="list-style-type: none"> • Stand der Forschung • Evaluationsworkflow • Grundmodell, bevor mit dem zeitaufwändigen Training und Parametertuning begonnen wird <p>Study Doc: Das Führen eines Study Docs soll ein systematisches, reproduzierbares Vorgehen, insb. für die Phase der Modellbildung, des Trainings und des Hyperparametertunings, erleichtern. Es enthält eine Problembeschreibung, einen Vorgehensplan, Ideen, Experimente, sowie deren Resultaten und Interpretationen. Für den Schlussbericht sollen idealerweise grosse Teile direkt aus dem Study Doc übernommen werden können.</p> <p>Zu verwendende Technologien:</p> <ul style="list-style-type: none"> • Muss: Git, Python, Tensorflow 2.0, Docker • Vorschlag: Python packages: librosa IDE: pycharm <p>Besprechungstermine: Besprechungstermine werden mit dem Betreuer individuell vereinbart und finden in der Regel alle zwei Wochen statt. Dazu ist dem Betreuer regelmässig am Vorabend einer Besprechung ein kurzer Status-Report (maximal 1/2 Seite) zukommen zu lassen, welcher stichwortartig über folgende drei Punkte Auskunft gibt:</p> <ul style="list-style-type: none"> • Welche Arbeiten wurden in der letzten Berichtsperiode ausgeführt, welche Arbeiten sind für die nächste Periode vorgesehen • Stand der Arbeiten (Soll- / Ist-Vergleich mit der Planung), ggf. Begründung von Abweichungen • Top-drei Risiken inkl. geplanter Massnahmen
Kreativität, Varianten, Innovation*	<ul style="list-style-type: none"> • Evaluation von verschiedenen Preprocessing Verfahren • Evaluation von unterschiedlichen Architekturen neuronaler Netze für das Embedding Modell. • Hyperparameter tuning
Schlagwörter:	Audiosignal, Neural Network, Maschinelles Lernen, Embedding, Geräusche Detektion
Wirtschaftsprojekt oder Bachelorarbeit:	<input type="checkbox"/> Wirtschaftsprojekt: 180 Stunden pro Studierender <input checked="" type="checkbox"/> Bachelorarbeit: 360 Stunden

* Bitte heben Sie in diesem Punkt hervor, inwiefern Ihre Projektidee **über kreativen Spielraum** verfügt. Dabei sind folgende Kriterien relevant: Die Idee erlaubt den Studierenden eigene Ideen zu entwickeln und Varianten zu erarbeiten, ist ausserhalb vom Tagesgeschäft angesiedelt, beinhaltet Neuland/Innovation und ist nicht durch Produkte & Tools getrieben.

Bitte kreuzen Sie eine Projektart und die zutreffenden Schwerpunkte an.

Projektarten:

- ☐ Einsatz von Standardsoftware und Services
- ☐ Software- und Produkt-Entwicklung
- ☒ Innovationsprojekte (Projekte mit Erkenntnisgewinn, Forschungsprojekte)
- ☐ IT-Infrastrukturentwicklung
- ☐ Strukturierte Analyse und Konzeption von Systemen und Abläufen

Schwerpunkte:

- ☒ Artificial Intelligence & Machine Learning
- ☐ Business Process Modelling
- ☐ Data Engineering
- ☐ Hardwarenahe Software-Erstellung
- ☐ Human Computer Interaction Design
- ☐ ICT Business Solutions
- ☐ ICT Infrastrukturen
- ☐ Internet of Things
- ☐ Mobile Systems
- ☐ Security/Privacy
- ☐ Software-Erstellung
- ☐ Visual Computing (Grafik, Bildverarbeitung, Vision, VR, AR)
- ☐ Anderes: _____

7. Zeiteinteilung

Vorschlag für die Zeiteinteilung pro Person

WIPRO:

pro Woche: ca. 12h
für Modulendprüfung: ca. 10h
Total: 180 h

BAA:

pro Woche: ca. 20h
Schlusswoche: ca. 50h
Für Modulendprüfung: ca. 10h
Total: 360 h

8. Rechtliche Grundlagen und Reglemente

Folgende Rechtsgrundlagen und Reglemente sind für die Wirtschaftsprojekte und Bachelorarbeiten an der Hochschule Luzern – Informatik massgebend:

- Studienordnung für die Ausbildung an der Hochschule Luzern, FH Zentralschweiz ([Link](#))
- Studienreglement für die Bachelor-Ausbildung an der Hochschule Luzern - Informatik ([Link](#))

9. Bestätigung

Mit der Kenntnisnahme der Aufgabenstellung bestätigen Student/in und Auftraggeber/in, dass

- Sie mit der Aufgabenstellung einverstanden sind.
- die Auftraggeberin/der Auftraggeber damit einverstanden ist, dass die Hochschule Luzern – Informatik für die Organisation einer Bachelorarbeit von ihr/ihm einen Kostenbeitrag von CHF 1'000.00 (inkl. MwSt.) pro Student/in erhebt. Dies gilt nicht für Arbeiten, welche berufsbegleitend Studierende in Verbindung mit ihrem Arbeitgeber/ihrer Arbeitgeberin machen und für HSLU interne Auftraggeber/innen. Für die Wirtschaftsprojekte wird kein Kostenbeitrag verrechnet.
- Betreuungspersonen und Experten uneingeschränkter Einblick in die Arbeit erhalten. Auch anlässlich von Präsentationen und Marketingaktivitäten kann die Arbeit der Öffentlichkeit gezeigt werden. Eine Zusammenfassung der Arbeit wird in jedem Fall veröffentlicht. Falls das Thema

vertraulich behandelt werden soll, muss der Aufgabenstellung eine entsprechende Vertraulichkeitserklärung beiliegen.

Datum: 23.01.2020

Die definitive Aufgabenstellung (pdf-Format) bitte per E-Mail an die Transferstelle senden, zwingend in Kopie an alle involvierten Parteien.

Anlaufstelle für alle Informationen im Zusammenhang mit studentischen Arbeiten sowie für Entgegennahme von Projektideen & Aufgabenstellungen:

Hochschule Luzern - Informatik
Transfer Services
Suurstoffi 1
6343 Rotkreuz
T: 041 228 24 66
E: transfer.informatik@hslu.ch