Univerza v Ljubljani
Skupni interdisciplinarni program druge stopnje
kognitivna znanost

v sodelovanju z Universität Wien, Univerzita Komenského v Bratislave in Eötvös Loránd Tudományegyetem

Aleš Žagar

Cross-lingual Approach to Abstractive Summarization

Medjezikovni pristop k abstraktivnemu povzemanju

magistrsko delo

Ljubljana, 2020

University of ljubljana
Middle european interdisciplinary master's
programme in cognitive science

in association with Universität Wien, Univerzita
Komenského v Bratislave and Eötvös Lloránd
Tudományegyetem

Aleš Žagar

Cross-lingual Approach to Abstractive Summarization

master's thesis

Supervisor: Prof. Dr. Marko Robnik Šikonja
Co-supervisor: Univ. Prof. Dr. Igor Farkaš

Ljubljana, 2020

Univerza v Ljubljani
Skupni interdisciplinarni program druge stopnje
kognitivna znanost

v sodelovanju z Universität Wien, Univerzita Komenského v Bratislave in Eötvös Loránd Tudományegyetem

Aleš Žagar

Medjezikovni pristop k abstraktivnemu povzemanju

magistrsko delo

Mentor: prof. dr. Marko Robnik Šikonja
Somentor: univ. prof. dr. Igor Farkaš

Ljubljana, 2020

Besedilo je oblikovano z urejevalnikom besedil LaTeX.

Acknowledgments

I would like to express my gratitude to my supervisor prof. dr. Marko Robnik Šikonja for a guidance and the always thorough reviews of my work.

I wish to acknowledge the help and useful suggestions provided by my co-supervisor prof. dr. Igor Farkaš.

I would like to offer my special thanks to Rok Zidarn for the template of the human evaluation and to all human evaluators.

Finally, I wish to thank my parents and friends for their support and encouragement throughout my study.

Aleš Žagar, 2020

"It is not events that disturb people, it is their judgements concerning them."

— Epictetus, Enchiridion

Contents

List of used acronmys

acronym	meaning
BERT	bidirectional encoder representations from transformers
CNN	convolutional neural network
LM	language model
LSTM	long short-term memory
MLP	multi-layer perceptron
NLP	natural language processing
NN	neural network
RL	reinforcement learning
RNN	recurrent neural networks
ROUGE	recall-oriented understudy for gisting evaluation
Seq2seq	sequence-to-sequence model

Povzetek

Naslov: Medjezikovni pristop k abstraktivnemu povzemanju

Avtomatsko povzemanje besedil označuje proces pridobivanja pomembnih informacij iz besedila in njihovo predstavitev v obliki povzetka. Pristopi k abstraktivnemu povzemanju so precej napredovali z uporabo globokih nevronskih mrež, a so rezultati še vedno lahko nezadovoljivi, kar še posebej velja za jezike brez velikih učnih množic. Pri mnogih nalogah obdelave naravnega jezika se medjezikovni prenosi kažejo za uspešne tudi pri jezikih, ki za rešitev problema nimajo ustrezno velikih učnih množic. Doslej še ni bilo poskusa medjezikovnega prenosa povzemanja zaradi neenostavne ponovne uporabe dekodirnika nevronskih modelov. V delu smo za povzemanje slovenskih novičarskih člankov uporabili vnaprej naučen model za povzemanje v angleškem jeziku, ki temelji na globokih nevronskih mrežah in arhitekturi zaporedje v zaporedje. Problem neustreznega dekodirnika smo rešili z uporabo dodatnega jezikovnega modela za generiranje besedil v ciljnem jeziku. Razvili smo pet modelov, ki se med seboj ločijo po številu učnih primerov. Rezultate smo evalvirali z avtomatsko in človeško evalvacijo. Z našim medjezikovnim modelom smo dosegli primerljiv rezultat z obstoječim abstraktivnim povzemalnikom za slovenski jezik. V delu obravnavamo tudi relevantne interdisciplinarne vidike.

Ključne besede

avtomatsko povzemanje, generiranje besedil, globoke nevronske mreže, jezikovni modeli, medjezikovne vložitve, abstraktivno povzemanje

Abstract

Title: Cross-lingual Approach to Abstractive Summarization

Automatic text summarization is a process of extracting important information from texts and presenting that information in the form of a summary. Abstractive summarization approaches progressed using deep neural networks, but results are not yet satisfactory, especially for languages where large training sets do not exist. In several natural language processing tasks, cross-lingual model transfers are succesfully applied for low-resource languages where large enough datasets are not available. For summarization such cross-lingual transfer was so far not attempted due to non-reusable decoder side of neural models. In our work, we used a pretrained English summarization model based on deep neural networks and sequence-to-sequece architecture to summarize Slovene news articles. We solved the problem with inadequate decoder by using an additional language model for target language text generation. We developed five models with different training sample sizes. The results were assessed by automatic and human evaluation. Our cross-lingual model performance is similar to the existing Slovene abstractive summarizer. We also discuss some interdisciplinary aspects, raised by our work.

Keywords

automatic summarization, text generation, deep neural networks, language models, cross-lingual embeddings, abstractive summarization

Razširjeni povzetek

Količina besedil v svetu hitro narašča. Motivacija za izgradnjo avtomatskih povzemalnih orodij je pospešitev seznanitve s pomembnimi informacijami in obvladovanje informacijske poplave. Takšna orodja se uporabljajo v različnih iskalnikih, v novinarstvu, pri sledenju novicam, poslovnih in pravnih analizah, medicinskih primerih ipd. Konkretne primere uporabe v praksi predstavljajo spletni povzemalniki člankov in besedil (SMMRY, 2019; Resoomer, 2019), raširitve spletnih brskalnikov (TLDR This, 2019), urejevalniki spletnih vsebin (Adobe, 2020) in povzemalniki zdravstvenih datotek (Hirsch et al., 2015).

Povzemalna orodja obstajajo le za jezike z zadostnimi zbirkami jezikovnih virov. Z razvojem medjezikovnih vektorskih vložitev je nastala priložnost za prenos naučenih modelov na tehnološko manj podprte jezike (Adams et al., 2017). Za povzemalnike zaradi tehnoloških in jezikovnih ovir takšen prenos še ni bil izveden. Predlagani pristop predstavlja prvi poskus prenosa naučenega abstraktivnega povzemalnega modela na drug jezik.

I Opredelitev problema

Povzemanje besedil je proces pridobivanja pomembnih informacij iz besedila in njihove predstavitve v obliki povzetka. Razdelimo ga lahko na abstraktivno in ekstraktivno povzemanje. Slednje kopira pomembne stavke iz originalnega besedila, medtem ko je abstraktivni pristop ustvarjalen in lahko v povzetek vključi tudi nove stavke in besede.

Pristopi abstraktivnega povzemanja zastavijo problem na podoben način kot strojno prevajanje, vendar s pomembnimi razlikami. Dolžina izhodnega besedila je bistveno krajša od vhodnega, povzetek pa mora izpustiti nepomembne informacije (Nallapati et al., 2016). Trenutni rezultati abstraktivnega povzemanja so lahko ponavljajoči se, absurdni, popačijo dejstva, nezadostno obravnavajo besede izven slovarja, slabo izbirajo relevantno vsebino in so lahko zavajajoči še v drugih ozirih. Kljub temu pa povzemalniki pogosto proizvedejo uporabne vsebine.

II Kratek pregled sorodnih del

Starejši povzemalni pristopi so bili povečini ekstraktivni, kar je omogočalo hkratno povzemanje večih dokumentov (Gambhir & Gupta, 2017). Ti pristopi se razvijajo tudi danes (Canhasi & Kononenko, 2016) predvsem s predstavitvijo v obliki grafov. Problema abstraktivnega povzemanja so se sprva lotevali z uporabo iterativnih algoritmov krajšanja stavka, kot je na primer izločanje posameznih besed (Knight & Marcu, 2002), kar je bilo kasneje razširjeno s substitucijo, menjavo vrstnega reda ali vstavljanjem novih besed (Cohn & Lapata, 2008).

Sodobni pristopi uporabljajo rekurenčne globoke nevronske mreže (Rush et al., 2015; Nallapati et al., 2016), ki najprej predstavijo izvorni dokument v numerični obliki in ga nato pretvorijo v povzetek. Takšni modeli so najuspešnejši za krajše povzetke, kot sta na primer povzemanje krajših novic ali generiranje časopisnih naslovov. Pomembna komponenta teh modelov je mehanizem pozornosti, ki zagotavlja, da se dekodirni del osredotoči na ustrezne vhodne besede. Nekateri modeli vključujejo tudi mehanizem kopiranja, ki omogoča, da lahko model vključi v povzetek besede, ki niso prisotne v slovarju, in mehanizem pokrivanja vsebin, ki preprečuje ponavljanje (See et al., 2017).

Zidarn (2019) je razvil prvi abstraktivni povzemalnik za slovenski jezik, ki temelji na rekurenčnih globokih mrežah. Najboljše rezultate je dosegel z

dvonivojsko LSTM nevronsko mrežo, mehanizmom pozornosti, mehanizmom kopiranja in iskanjem v snopu. Sprva je poskusil povzemalnik naučiti na slovenski Wikipediji, ki pa se je izkazala za premajhno in neustrezno, zato je uporabil novice Slovenske tiskovne agencije (STA).

Nevronske mreže za obdelavo besedil ta najprej pretvorijo v vektorsko obliko, s postopkom, ki ga imenujemo vektorske vložitve. Ideja besednih vložitev je, da besede predstavimo kot visoko dimenzionalne vektorje, ki predstavljajo pomen besed. Znani različici teh vložitev sta Word2vec (Mikolov, Chen, et al., 2013) in fastText (Grave et al., 2018). Slabost teh vložitev je neupoštevanje večpomenskosti besed. Novejše različice vektorskih vložitev, kot sta ELMo (Peters et al., 2018) in BERT (Devlin et al., 2019), pri tvorbi v vektorsko obliko upoštevajo kontekst besed in generirajo numerično predstavitev odvisno od konteksta, zato ima vsaka beseda mnogo različnih predstavitev.

Pomemben uvid za naše delo je, da se relacije med besedami v vložitvenih prostorih jezika ohranjajo (Mikolov, Le, & Sutskever, 2013). Z različnimi tehnikami lahko poravnamo enojezične vložitve v skupni vektorski prostor, s čimer dobimo medjezikovne vložitve (Ruder et al., 2019). Na začetku so te tehnike zahtevale paralelni korpus in/ali dvojezični slovar, ki je zagotavljal potrebne informacije o tem, kateri besedi v izvornem in ciljnem jeziku je potrebno poravnati. Z novejšimi pristopi (Conneau et al., 2017) je možno medjezikovne vložitve proizvesti tudi z nenadzorovanim učenjem brez dvojezičnih virov.

III Interdisciplinarni vidiki

Pomembno vprašanje za naše delo je, kako beseda pridobi pomen. V splošnem ločimo semantične in referenčne teorije pomena. Semantične teorije se ukvarjajo s pojasnjevanjem informacij, ki izhajajo iz naravnega jezika in vplivajo na pomen besede, referenčne teorije pomena pa se ukvarjajo s pojasnjevanjem dejstev in stanj stvari, na podlagi katerih besede pridobijo pomen (Gasparri & Marconi, 2019). V raziskavah besednih vložitev je definicija pomena be-

sede vzeta iz distribucijske semantike: "pomen besede prepoznaš po besedah, s katerimi je ta v družbi"(Firth, 1957). Izvor takšnega pojmovanja pomena najdemo v *Filozofskih raziskavah* Ludwiga Wittgenteina (2011), ki definira pomen besede kot njeno uporabo v jeziku. Vendar pa pomen nekaterih besed definiramo tudi s kazanjem na referenta oz. označenca. Takšno pojmovanje izpostavi dejstvo, da je pomen besede sestavljen tako iz naravnega jezika kot tudi iz stanja stvari oz. dejstev, vendar v različnih deležih. Na tej podlagi ne moremo reči, da lahko pomen besede izpeljemo samo na podlagi njenega pojavljanja. Kljub temu pa raziskave besednih vložitev kažejo na to, da je precejšen napredek možen samo z upoštevanjem pojavljanja besed. To podpira hipotezo, da je vsaj del besednega pomena sestavljen iz njenega pojavljanja. V tem vidimo priložnost za sodelovanje med področjema obdelave naravnega jezika in konceptualne analize pomena. Raziskave obdelave naravnega jezika nam lahko predstavijo, v kakšnem deležu je pomen določen s pojavitvami, konceptualna analiza pa razstavi glavne komponente besednega pomena.

Naslednji interdisciplinarni vidik sestavljata vprašanje obstoja in razlage jezikovnih univerzalij. Jezikovne univerzalije so skupne poteze naravnih jezikov. Sintaktična razlika med glagolom in samostalnikom je na primer ena izmed takšnih potez. Cowie (2017) povzame številne razlage jezikovnih univerzalij, ki so lahko posledica vrojene splošne slovnice (Chomsky, 1987), izpeljane iz splošnih potreb sporočanja in opisovanja dejstev (Sapir, 2004), ali pa so se vsi jeziki razvili iz enega skupnega jezika (Cavalli-Sforza, 2001). Menimo, da lahko skupni vektorski prostor medjezikovnih vložitev besed razumemo kot eno izmed jezikovnih univerzalij. S tem dodatno podpremo obstoj takšnih univerzalij in morda tudi kakšno izmed tekmujočih razlag nastanka.

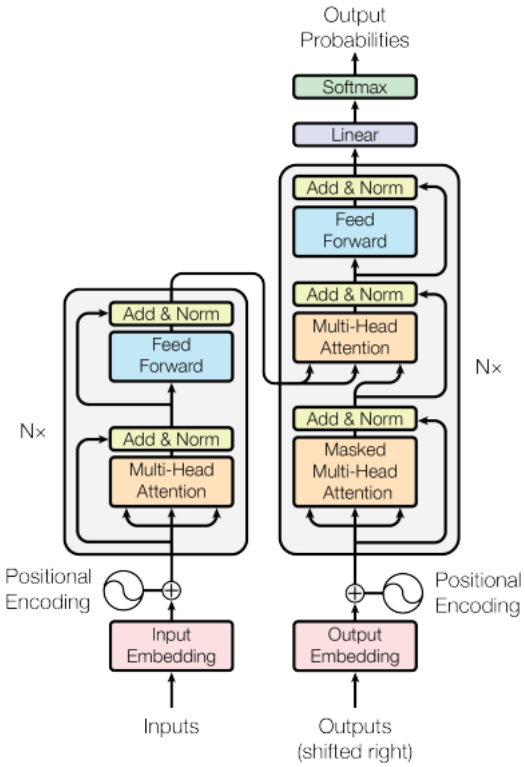
Tretji interdiscipinarni vidik predstavlja opažanje, da so sistemi umetne inteligence napredovali v reševanju številnih problemov (na primer strojno prevajanje, računalniški vid ipd.), ne da bi bilo v ta napredek vključeno tudi tako zaželeno razumevanje problemov, ki jih sistem rešuje. Na začetku raziskav umetne inteligence se je na primer nekritično antropomorfiziralo sisteme

kot "misleče stroje", kar je še danes prevladujoče razumevanje v širši javnosti, v znanosti pa so se raziskave obrnile predvsem v reševanje praktičnih problemov. Ne glede na to še vedno obstaja pomembna razlika med razumevajočim človeškim reševanjem problemov in nerazumevajočim računskim pristopom umetne inteligence. Glede tega želimo izpostaviti nevronske pristope, ki zavračajo predpostavko, da so simbolne reprezentacije in psihološki procesi na splošno nujni za uspešno izvajanje jezikovnih nalog, kot sta na primer prevajanje in povzemanje. Opozoriti želimo predvsem na to, da ne poznamo meja pristopov, ki ne uporabljajo simbolnih reprezentacij. Iz tovrstnih opažanj lahko izpeljemo vsaj dva sklepa. Psihološki proces, kot je na primer mišljenje, lahko, prvič, interpretiramo kot posebno človeško hevristiko in ne kot edini nujni način, kako uvideti bistvo problema, in drugič, ti procesi so lahko samo zavestna, subjektivna izkušnja računske narave človeškega mišljenja. Takšna zastavitev približa človeka stroju in ne obratno ter s tem spreminja njegovo samorazumevanje.

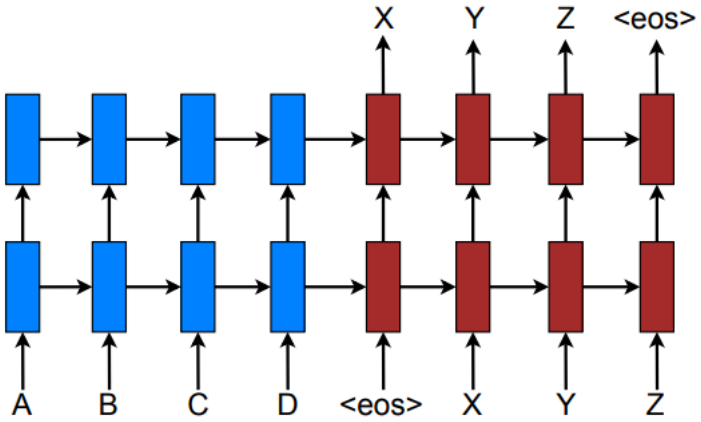
IV Metodologija in rezultati

V delu smo najprej ustvarili učni množici za povzemalnik in jezikovni model. STA učna množica je sestavljena iz parov povzetek-besedilo, pri čemer povzetek predstavlja uvodni odstavek izvornega besedila Slovenske tiskovne agencije (STA, 2019). Učno množico za jezikovni model (arhitektura modela je prikazana na sliki 1) predstavlja korpus digitalne slovenščine Gigafida (*Gigafida 2.0*, 2019).

Chen and Bansal (2018) sta avtorja vnaprej naučenega angleškega povzemalnika, ki smo ga uporabili. Gre za hibridni povzemalnik, ki najprej izvleče informativne stavke in jih za tem okrajša. Oba koraka, ekstraktivni in abstraktivni, sta naučena ločeno, sestavljajo ju različne nevronske mreže. Konvolucijska mreža predstavi stavek, prva rekurentna mreža kodira odvisnosti med stavki in druga rekurentna mreža napove, katere stavke bo model izvlekel. Abstraktivni korak je sestavljen iz rekurentnih nevronskih mrež in



Slika 1: Transformer arhitektura. Levo je predstavljen kodirnik, desno je predstavljen dekodirnik. Za naš jezikovni model smo uporabili zgolj dekodirnik, ki nam izračuna verjetnost za neko zaporedje znakov. Model temelji izključno na različnih mehanizmih pozornosti. Slika vzeta iz Vaswani et al. (2017).



Slika 2: Po zaporedju razvit prikaz dvonivojske kodirnik-dekodirnik arhitekture, sestavljene iz dveh rekurenčnih nevronskih mrež. Modro obarvane celice ponazarjajo kodirnik, ki izračuna (kodira) predstavitev vhodnih podatkov v obliki vektorjev. Rdeče obarvane celice ponazarjajo dekodirnik, ki uporabi te vektorje za napovedovanje primernih besed. Na sliki poskuša mreža iz zaporedja A, B, C, D napovedati X, Y, Z. $< eos >$ je poseben simbol, pri katerem mreža preneha tvoriti stavek ali besedilo. Slika vzeta iz Luong et al. (2015).

standardno kodirnik-dekodirnik arhitekturo (glej sliko 2) z mehanizmom kopiranja. Model je v zadnjem koraku optimiziran s spodbujevalnim učenjem.

Iz vnaprej naučenega angleškega modela smo izluščili angleške besedne vložitve. Vnaprej naučene slovenske besedne vložitve fastText (Grave et al., 2018) smo poravnali v jezikovni prostor angleških besednih vložitev s prokrustovsko poravnavo (Conneau et al., 2017).

Na angleškem modelu smo zamenjali angleške vložitve s slovenskimi in tako dobili osnovni povzemalni model. Osnovnemu modelu, ki ne vsebuje slovenskih učnih primerov, smo postopoma dodajali slovenske učne primere. Na ta način smo dobili štiri modele z različnimi razmerji učnih primerov. Poleg opisanih modelov smo naučili tudi izključno slovenski model. Rezultati neoptimiziranih modelov so prikazani v tabeli 1. Rezultati ponazarjajo, da je

Model	ROUGE-1	ROUGE-2	ROUGE-L
MENG	18,91	3,74	16,27
M1	12,94	1,96	11,61
M10	15,71	3,71	13,87
M100	**21,67**	**6,81**	**19,16**
MSLO	21,07	6,62	18,64

Tabela 1: Vrednosti ROUGE neoptimiziranih medjezikovnih abstraktivnih modelov povzemanja na slovenščini. Mera ROUGE meri prekrivanje vsebine izvornega referenčnega povzetka in generiranega povzetka modela. Višja vrednost pomeni boljše prekrivanje in s tem boljši povzetek. Model MENG ne vsebuje nobenega slovenskega učnega primera. Sledijo modeli z 1%, 10% in 100% deležem razpoložljivih slovenskih učnih primerov. Model MSLO ni medjezikoven in je naučen izključno na slovenski učni množici.

Parameter 1	ROUGE-1	ROUGE-2	ROUGE-L
Baseline M100	21,67	6,81	19,16
Transformer	22,53	6,83	19,61
BERT	24,87	**7,41**	21,36
ROUGE-L	**24,88**	7,38	**21,47**

Tabela 2: Otimizacija najboljšega primera z enim parametrom.

medjezikovni model boljši od modela naučenega izključno na slovenski učni množici za 0,60 točke pri ROUGE-1, 0,19 točke pri ROUGE-2 in 0,52 pri ROUGE-L. Glede na ta rezultat je medjezikovni prenos upravičen. Primeri povzetkov se nahajajo v dodatku A.

Sledila je optimizacija najboljšega modela. Iz vsakega modela smo vzeli množico hipotez, ki smo jih dodatno ovrednotili z ločeno naučenim jezikovnim modelom, vrednostjo BERTScore in vrednostjo ROUGE. Vrednosti BERTscore in ROUGE smo pridobili tako, da smo primerjali podobnost med izvornim izvlečenim stavkom in generiranim stavkom povzemalnika. Rezultati

Parameter 1	Parameter 2	ROUGE-1	ROUGE-2	ROUGE-L
ROUGE-L	BERTScore	24,97	7,43	21,50

Tabela 3: ROUGE vrednosti končnega modela, optimiziranega z dvema parametroma.

Model	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore
Zidarn (2019)	23,77	7,97	23,95	0,679
Žagar (2020)	24,97	7,43	21,50	0,679
Chen and Bansal (2018)	40,88	17,80	38,54	\
J. Zhang et al. (2019)	44,17	21,47	41,11	\

Tabela 4: Primerjava rezultatov našega povzemalnega modela s slovenskim (Zidarn, 2019) in dvema angleškima modeloma. Spomnimo, da smo vnaprej naučen angleški model iz tretje vrstice (Chen & Bansal, 2018) uporabili v naši raziskavi. Prilagamo tudi rezultate trenutno najuspešnejšega povzemalnika.

so prikazani v tabeli 2. Ločeno naučen jezikovni model Transformer izboljša model Baseline M100 za 0,86 točke pri ROUGE-1, 0,02 točke pri ROUGE-2 in 0,45 točke pri ROUGE-L. Rezultat pomembno izboljšata vsebinski metriki.

Odločili smo se, da množico hipotez poskusimo zamejiti tudi z dvema parametroma. Jezikovni model in interna ocena povzemalnika naj bi bila korelirana z berljivostjo hipoteze, vrednosti BERTScore in ROUGE pa vsaj nekoliko korelirana s prekrivanjem vsebine. Rezultat najboljšega modela je prikazan v tabeli 3. Izkazalo se je, da najboljša parametra upoštevata le vsebinsko prekrivanje. Razlog je verjetno v tem, da je končna ROUGE evalvacija povzemalnega modela vsebinska. Ročni pregled primerov je pokazal, da so modeli z mešanimi kombinacijami parametrov (eden za berljivost in eden za vsebinsko prekrivanje) na vtis nekoliko boljši, kar ponovno izpostavi pomanjkljivosti avtomatskih metrik za ocenjevanje povzetkov. Primeri povzetkov najboljšega modela se nahajajo v razdelku 6.4.2.

Tabela 4 ponazarja primerjavo našega modela s slovenskim modelom

in angleškima modeloma. Poleg standardne ROUGE evalvacije povzetkov poročamo tudi vrednosti BERTScore. V primerjavi s slovenskim modelom so naši rezultati višji pri ROUGE-1 za 1,20 ter nižji pri ROUGE-2 in ROUGE-L za 0,54 in 2,45 točke. Vrednosti BERTScore sta identični. Človeška evalvacija je pokazala, da naš model proizvede več točne vsebine, oba pa proizvedeta povzetke s sprejemljivo berljivostjo. Glede na to, da smo uporabili različne podmnožice izvornih novic STA, različno razdelitev na učne in testne podatke in glede na problematično naravo avtomatskih mer ocenjevanja povzetkov, lahko zaključimo, da sta modela po učinkovitosti med seboj podobna, vendar pa še ne moremo govoriti o njuni praktični uporabnosti.

Slovenski modeli se ne morejo primerjati z angleškimi, saj so ti naučeni na številčnejših in bolj ustreznih učnih podatkih za povzemanje besedil. Dosegajo skoraj dvakrat višje vrednosti primerjanih mer kakovosti povzemanja od naših. Povzetki so točnejši in berljivejši.

V Sklep

V nalogi smo razvili prvi medjezikovni nevronski model za abstraktivno povzemanje besedil. Preverili smo, kako dodajanje učnih primerov vpliva na učinkovitost modela. Naučili smo jezikovni model, s katerim smo poskušali olajšati težave medjezikovnega prenosa. Poleg strojne evalvacije proizvedenih povzetkov smo uporabili tudi človeško evalvacijo. Ustvarili smo učno množico novic STA, primerno za strojno povzemanje besedil.

Naš medjezikovni pristop proizvede povzetke tudi z malo podatki v ciljnem jeziku. Z veliko množico podatkov v ciljnem jeziku je uspešnost primerljiva z modelom, naučenim izključno v ciljnem jeziku. Rezultati potrjujejo, da je uspešnost nevronskih mrež odvisna od količine podatkov. Naš povzemalnik uspešneje povzema bolj zastopane teme v učni množici (politične in finančne novice). Človeška evalvacija kaže, da povzemalnik proizvede povzetke z zadovoljivo točnostjo in s sprejemljivo berljivostjo.

Model bi lahko izboljšali z uspešnejšo medjezikovno poravnavo vložitev

ali kontekstualnimi vložitvami. Zaradi kompleksne morfologije slovenskega jezika bi bilo smiselno povečati število besed v slovarju povzemalnika. Namesto mere ROUGE bi kot nagrado pri optimizaciji modela s spodbujevalnim učenjem lahko kot nagrado uporabili mero BERTScore. Berljivost povzetkov bi lahko ovrednotili z modeli strojnega učenja za ocenjevanje berljivosti. Namesto hevristike bi lahko priskrbeli izvorne razdelitve člankov v uvodni odstavek in telo besedila. Iz učne množice bi lahko na podlagi vrednosti BERTScore izločili neustrezne uvodne odstavke.

Sedanje mere za oceno abstraktivnih povzetkov so koleriranje le z vsebino povzetkov. V prihodnjih študijah bi se lahko lotili izboljšav teh mer z upoštevanjem berljivosti in vsebinske točnosti generiranih povzetkov. Zanimiv problem predstavlja vprašanje, kako povečati abstraktivnost povzetkov, saj se najboljši modeli dobro naučijo izpuščati manj pomembne besede in stavčne strukture, slabše pa jim gre ustvarjanje novih besed in zamenjava daljših stavčni struktur s krajšimi. Pri medjezikovnem prenosu modelov, ki vključujejo generiranje besedil, se zastavlja problem poravnave posebnih značk, kot je na primer značka za ustavitev generiranja besedila.

Chapter 1

Introduction

The amount of text data in the world is rapidly growing. Automatic summarization tools can help us gain important information in less time. These tools are used in search engines, for business analyses, medical cases, legal documents, etc. Some concrete examples include web-based articles and texts summarizers (SMMRY, 2019; Resoomer, 2019), browser extensions (TLDR This, 2019), web content management systems (Adobe, 2020) and patient record summarizers (Hirsch et al., 2015).

In future, automatic text summarizers will become part of personal assistants, question-answering systems, and automated content creation. Possible applications include monitoring trends and innovations within scientific research, helping people with hearing disabilities by condensing voice-to-text technology (Ratia, 2018), etc.

1.1 Automatic summarization problem

Summarization is a process of extracting or collecting important information from texts and presenting that information in the form of a summary. It is broadly divided into extractive and abstractive summarization. The extractive approach is non-productive in a sense that it just copies important sentences, and the resulting summary does not include any new words

or sentences. The abstractive approach is creative and produces summaries that rephrase the given content in compressed sentences that can contain originally unused words.

The abstractive neural summarization task uses similar deep learning approaches but is different from machine translation. The output of summarization is short compared to the input, and the compression is lossy (Nallapati et al., 2016). Current abstractive summarization approaches can be repetitive, absurd, can misrepresent facts, deal poorly with out-of-vocabulary words, perform poorly at content selection, and can be misleading in various other ways. Nevertheless, they often produce useful outputs.

Many summarization tools exist for resource-rich languages. Cross-lingual embeddings present a promising approach for low resource languages and enable the model transfer from resource-rich to resource-poor languages (Adams et al., 2017; Artetxe & Schwenk, 2019). It is an open problem whether the cross-lingual approach is feasible for abstractive summarization and our work presents the first attempt to transfer trained summarization model to another language.

Our goal is to develop a cross-lingual abstractive summarization approach. Our main research questions are: 1. What is the performance of the summarization model trained with word embeddings aligned in a common vector space? 2. Are cross-lingual text summarization models competitive to monolingual models?

1.2 Related research

Most of early summarization approaches use the extractive approach which is also suitable for a multi-document summarization (Gambhir & Gupta, 2017) and is still evolving (Canhasi & Kononenko, 2016). Some authors modeled the abstractive summarization problem as the sentence compression task. The very first ones used a word deletion technique (Knight & Marcu, 2002) which was later extended to word substitution, reordering and insertion

(Cohn & Lapata, 2008).

In addition to the abstractive and extractive text summarization, the approaches can be categorized based on an intermediate representation of the text. Allahyari et al. (2017) distinguish topic representation approaches that interpret the topics discussed in the text, and indicator representation approaches that represent sentences as a list of features. Examples of the former use topic words or word probabilities to determine the sentence importance. The latter approaches construct graphs that group similar sentences and select the important ones, or machine learning techniques that divide sentences into summary and non-summary classes and treat the problem as a classification task. Other summarization approaches make use of knowledge bases or exploit the context of summarization (web pages, scientific articles, emails).

Lately, deep neural networks learning sequence to sequence (seq2seq) transformations produced state of the art abstractive summaries (Rush et al., 2015; Nallapati et al., 2016). They encode a source document into an internal numeric representation and then decode it into an abstractive summary. These models work best for a short single-document summaries, e.g., headline generation, news summarization, and we will use them in our approach. In general, they use the attention mechanism which ensures that the decoder focuses on the appropriate input words. Some additional mechanisms are the copy mechanism that copies relevant words from the input when they are not present in a dictionary (See et al., 2017), and the coverage mechanism that avoids the repeating content (Tu et al., 2016).

Recently, Zidarn (2019) built the first abstractive summarizer for the Slovene language using the seq2seq architecture and deep neural networks. The best results were produced by a two-layer LSTM with attention mechanism, copy mechanism, and beam search. Initially, the Slovene Wikipedia dataset was considered but it was too small, so the STA news dataset (STA, 2019) was used instead.

The idea of word embeddings is to learn high-dimensional vectors that capture the meaning of words. Popular variants are Word2vec (Mikolov,

Chen, et al., 2013), GloVe (Pennington et al., 2014), fastText (Grave et al., 2018), ELMo (Peters et al., 2018), and BERT (Devlin et al., 2019).

An important insight for this work is that relations between words in the embedded space are preserved across the languages (Mikolov, Le, & Sutskever, 2013). Cross-lingual embeddings align monolingual embeddings into a common vector space (Ruder et al., 2019). In the beginning, these techniques required parallel corpora or a bilingual dictionary that provide necessary information for mapping a word from a source to a target language. The recent approach can train cross-lingual embeddings in an unsupervised manner (Conneau et al., 2017).

A major drawback of these classical word embeddings is that they cannot deal with polysemy. Recent contextual embeddings BERT (Devlin et al., 2019) and ELMo (Peters et al., 2018) learn polysemic representations of words.

1.3 Interdisciplinary aspects

In this section, we first briefly present a general problem of interdisciplinary research and then address a few research topics of cognitive science addressed in our work.

Firstly, we want to emphasize that there are many incommensurabilities between the fields involved in our thesis. To prove or disprove the theories or hypotheses traversing diverse scientific paradigms takes a substantial work, and may not be possible without interpreting the fundamental beliefs of each discipline (Kuhn, 1998). One has to provide a concise representation of all related concepts to make any meaningful conclusions. In some cases even that is not enough. Sometimes superficial similarities conceal deeper methodological incompatibilities, and sometimes it is illusory that the work even addresses the same topic. Different areas need different methodologies, which are comprised of different theories of justification. Taking that into consideration, we are aware that our discussion may be flawed.

Secondly, the addressed questions are general while our research is specific. No single research, similar to ours, can answer the questions we present in the following subsection. We believe that only the whole fields of research (e.g. natural language processing) can contribute possible answers.

1.3.1 Interdisciplinary research aspects

One of the interdisciplinary aspect of our work is the question how words acquire meaning. In general, theories of word meaning are divided into semantic and foundational theories. A semantic theory "is a theory interested in clarifying what meaning-determining information is encoded by the words of a natural language", and a foundational theory of word meaning "is a theory interested in elucidating the facts in virtue of which words come to have the semantic properties they have for their users" (Gasparri & Marconi, 2019). In word embeddings research, word meaning is stated as "you shall know a word by the company it keeps", which is a distributional semantics' hypothesis and attributed to Firth (1957). But the idea originates from Wittgenstein's *Philosophical Investigations*, where he defines the meaning of a word as its use in the language: "For a large class of cases of the employment of the word "meaning" - though not for all - this word can be explained in this way: the meaning of a word is its use in the language. And the meaning of a name is sometimes explained by pointing to its bearer" (2011). The quote shows how both theories of word meaning overlap, but in different proportions, since no theory of word meaning is strictly semantical or foundational. According to this account, words usually get their meaning from the contexts in which they are used, and only secondarily word meanings are explained by pointing to a bearer, a referent, a fact, or a state of affairs. In this sense, the meaning can be derived neither purely statistically nor not statistically at all. Word embeddings research shows that the progress is possible by taking into account word occurrences only. This supports the hypothesis that at least part of the meaning can be explained statistically. Natural language processing may in the future help us determine specifically

how much of word meaning can be explained purely statistically, and other fields can through the conceptual analysis explicate the principal components of the concept.

Another aspect deals with the problems of the existence and explanation of universals across languages. Linguistic universals are features thought to be common to all natural languages[1], e.g., the existence of a syntactic distinction between nouns and verbs (Cowie, 2017). Many explanations of universals exist (Cowie, 2017). They could be consequences of speakers' innate knowledge of universal grammar (Chomsky, 1987), could derive from universal demands of the communication situation - language is used to communicate propositions (Sapir, 2004), or suggest that all human languages have probably evolved from the language spoken by a small group of humans migrating from Africa (Cavalli-Sforza, 2001). With further research, a common vector space, studied in cross-lingual word embeddings research, can be interpreted as a linguistic universal and may, firstly, additionally support the existence of the linguistic universals, and secondly, support one of the competing explanations.

The third interdisciplinary aspect constitutes an interesting fact that the artificial intelligence systems have greatly improved over the years (e.g. language translation, computer vision, etc.) without a desirable improvement in the comprehension or understanding of the problems. We consider such an improvement as an unjustified anthropomorphization, e.g. the concept of "thinking machines" in the early stages of AI research. However, there still remains an essential distinction of problem solving abilities between humans and machines. We want to indicate that more intriguing than the question "whether machines can think?" are neural approaches that seem to refute the assumption that symbolic representations or psychological processes in general are necessary conditions for solving natural language tasks such as translation or summarization. In this sense, we want to emphasize how much can be achieved without symbolic representations of the contents and raise a

[1]Greenberg (1963) identified 45 linguistic universals across 30 studied languages.

question whether AI will hit the wall because of that. Psychological processes (e.g. thinking) may be interpreted, firstly, only as a specific human heuristic and not the single way to grasp the essence of a problem, and secondly, these processes could be considered as conscious subjective experiences of one's own computational nature of thinking processes. This shifts questions in the direction of making humans more like machines and not vice versa.

1.4 Structure

The remaining chapters are organized as follows. In Chapter 2, we present technologies and methods used. In Chapter 3, we describe the process of building a training dataset. In Chapter 4, we present the training, the solution scheme and the architectural choices. In Chapter 5, we deal with evaluation metrics. In Chapter 6, we present results and discussion. In Chapter 7, we conclude and present ideas for future work.

Chapter 2

Methodology

In this chapter, we introduce the components of our approach to cross-lingual summarization: deep neural networks, language models, and word embeddings. We explain the specific architectures and models used in our work.

2.1 Deep neural networks

The basic unit of a neural network is an idealized artificial neuron, inspired by the biological neurons in the brain, and conceived by McCulloch and Pitts (1943). The unit represents the (non-linear) transformation of the weighted sum of its inputs. It is illustrated in Figure 2.1 and defined by:

$$y = f\big(\sum_{i=1}^{n} w_i x_i + b\big) \tag{2.1}$$

The weights w_i represent the strength of connections and indicate the importance of a particular input. The activation function f is a non-linear transformation of the linear output and gives the neuron more power to accurately model complex problems. Without the activation function, the unit is just a linear regression model. Many activation functions were proposed (sigmoid, ReLU, tanh, etc.) with different strengths and weaknesses. The question which activation function to choose depends on the problem and

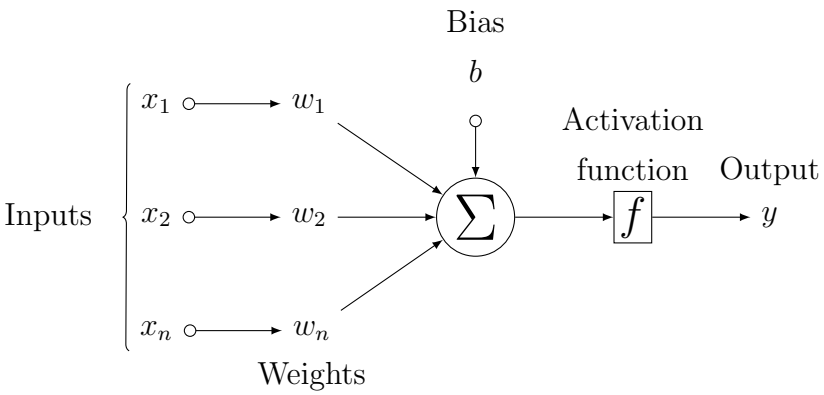


Figure 2.1: Artificial neuron

is often resolved through the model optimization. The most widely used activation function in text processing is the ReLU function.

One artificial neuron or a perceptron is not much of a practical use. From Figure 2.2 we can see that a basic neural network consists of several neurons and layers, and is called a multi-layer perceptron (MLP). Layers are usually fully connected which means that each node gets its input from all of neurons in the previous layer. The number of hidden layers is a chosen hyperparameter and is not restricted to one. Different layers can have different activation functions. The equations for a network with two hidden layers would be:

$$h_1 = f\big(W_1 x + b_1\big) \tag{2.2}$$

$$h_2 = f\big(W_2 h_1 + b_2\big) \tag{2.3}$$

$$y = f\big(W_{out} h_2 + b_{out}\big), \tag{2.4}$$

where W_1, W_2 and W_out are weight matrices, b_1, b_2 and b_out are bias terms and h_1 and h_2 are outputs of hidden layers 1 and 2. The weight matrices and bias terms are the parameters to be optimized during training. The goal of training is to minimize the error function or more generally the loss function. There are many loss functions, e.g., mean squared error for regression problems and cross-entropy loss for classification problems.

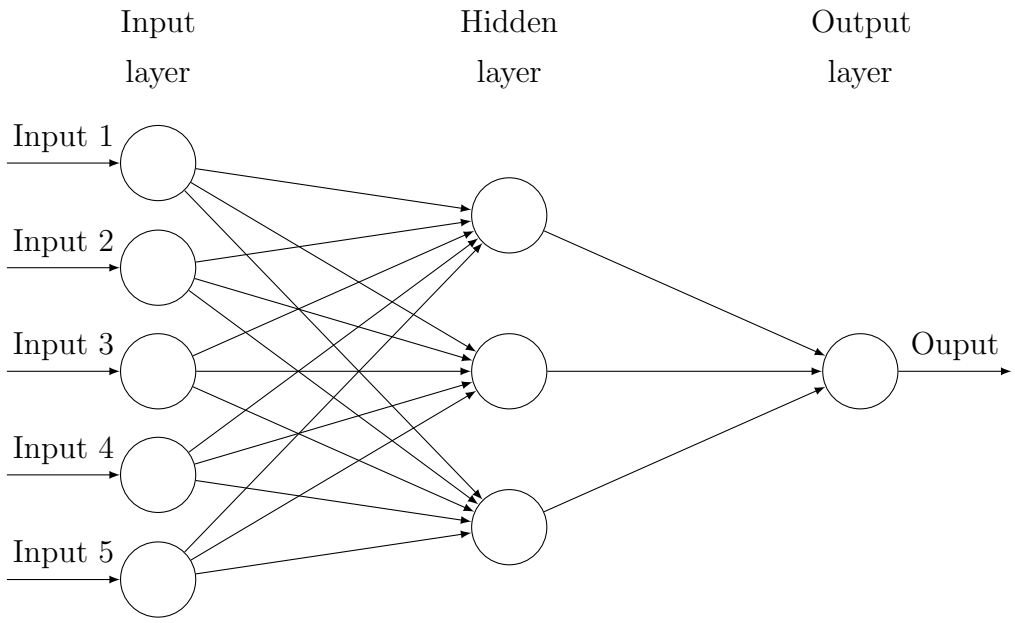


Figure 2.2: Multi-layer perceptron

At the beginning of training, the parameters are randomly initialized. Many different weight initialization techniques were proposed, from a random normal distribution to more advanced ones like He (He et al., 2015) or Xavier (Glorot & Bengio, 2010)). The parameters are usually initialized to small random values around 0. If the weights in a network are too small, then the error gradient shrinks as it back propagates through each layer until it is too tiny to be useful (vanishing gradient). If the initial weights in a network are too large, then the signal grows as it passes through each layer until it's too massive to be useful (explosive gradient).

A standard approach for training neural networks is the backpropagation algorithm (Werbos & John, 1974), although alternative approaches exist (Malinovská et al., 2018). Backpropagation was rediscovered several times but was finally credited to Rumelhart, Hinton & Williams (1986). The backpropagation computes the gradient of the loss function with respect to each parameter. If we know the gradient, we can use the chain rule to find out

how the weights have to change to minimize the cost function. With this technique, we can arrive at the local minimum.

2.1.1 Recurrent neural networks

One of the weaknesses of MLP is that it cannot take into account the sequential nature of problems. In language, the order of words is an essential information. Recurrent neural networks[1] try to address this problem with an extension of MLP considering previous hidden states:

$$h_t = f_1(W_{(hh)}h_{t-1} + W_{(hx)}x_t), \tag{2.5}$$

$$\hat{y}_t = f_2(W_{(ho)}h_t), \tag{2.6}$$

where t is a timestep, h_{t-1} is the previous hidden state, h_t is the current hidden state, $W_{(hh)}$ is the weight matrix that transforms the previous hidden state, $W_{(hx)}$ is the weight matrix that transforms the input (usually a word vector), $W_{(ho)}$ is the weight matrix that transforms the current hidden state, and $\hat{y}_t$ is the predicted output, and f_1 and f_2 are the activation functions. Usually, biases are present but are omitted here for simplicity.

Basic RNNs suffer from the vanishing gradient problem when processing longer sequences. This problem is successfully tackled by long short-term memory networks (LSTMs). In LSTMs, at every step t, there is a hidden state $h^{(t)}$ and a cell state $c^{(t)}$. The LSTM can erase, write and read information from that cell which stores long-term information. Three gates control how the information flows through the cell. The forget gate f controls what is kept from the previous cell, the input gate i controls what parts of the new cell content are written to the cell and the output gate o controls what parts of the cell are transmitted to the hidden state. The new cell content $\tilde{c}$ contains information to be written to the cell c. The forget gate is applied using element-wise product $\circ$ on previous cell state to which the input from

[1] We will use the description and notation for RNNs and LSTMs from Mohammadi (2017).

the new cell content is added. Finally, the hidden state h reads some content from the cell. The weight matrices W and U are learning parameters that transform the previous hidden states and the inputs. The whole process is given by the equations:

$$f^{(t)} = \sigma(W_f h^{(t-1)} + U_f x^{(t)} + b_f) \tag{2.7}$$

$$i^{(t)} = \sigma(W_i h^{(t-1)} + U_i x^{(t)} + b_i) \tag{2.8}$$

$$o^{(t)} = \sigma(W_o h^{(t-1)} + U_o x^{(t)} + b_o) \tag{2.9}$$

$$\tilde{c}^{(t)} = tanh(W_c h^{(t-1)} + U_c x^{(t)} + b_c) \tag{2.10}$$

$$c^{(t)} = f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t)} \tag{2.11}$$

$$h^{(t)} = o^{(t)} \circ tanh(c^{(t)}). \tag{2.12}$$

$$\tag{2.13}$$

The next important upgrade of the RNNs is the attention mechanism. Understanding the attention requires understanding what is a sequence-to-sequence model. This model maps one sequence of tokens to another sequence of tokens. It is also known as an encoder-decoder architecture and is widely used in machine translation and abstractive text summarization models. A typical implementation uses a multi-layered LSTM to encode the input sequence to a vector of fixed dimensionality (also known as a context vector) and another LSTM that decodes the vector into the output sequence (Sutskever et al., 2014).

Bahdanau et al. (2015) assumed that only one such context vector, no matter the dimension, probably cannot process the input adequately and presents the information bottleneck of the model. This is especially true in text summarization task, where the inputs are much longer in comparison to machine translation. The solution to this problem is the attention mechanism that allows the model to attend to the relevant parts of the input sequence. In essence, the attention is additional memory devoted to relevant source words. With the attention mechanism, the encoder does not pass only one final vector to the decoder but all of its hidden states. The new context

vector is calculated by multiplying attention scores with hidden states. Thus, a model with attention can process long range dependencies.

Another important component of sequence-to-sequence models is the beam search. When the decoder is generating words, it is not always the best idea to take the word with the highest predicted probability in the output sequence. A better way is to store a set (i.e., a beam) of n best hypotheses (sequences) and monitor how they are evolving. It may happen that initially inferior hypothesis becomes the chosen one. The beam search can include many advanced features that regularize scores. A length normalization, for example, inhibits favoring short hypotheses, and a coverage mechanism encourages the decoder to attend to all the words in the input.

2.1.2 Transformer archictecture

In this section, we describe the network architecture we used for language modeling task. A transformer is an encoder-decoder architecture eschewing recurrence and relying completely on the attention mechanism (Vaswani et al., 2017). We provide the architecture in Fig. 2.3 and the attention in Fig. 2.4 from the original paper to help track the computations.

The encoder and decoder are composed of a stack of layers. Each layer consists of two sub-layers: a multi-head self-attention mechanism and a simple feed-forward network. A residual connection around each of the two sub-layers is employed, followed by a layer normalization.

Information flows through the transformer following the arrows. Since there is no recurrence, the model must encode the position of embeddings to make use of the order of the sequence. This is achieved by positional encoding where the transformer adds a position vector to each input embedding. The original implementation uses sine and cosine functions of different frequencies to encode positions but there are many alternatives, learned and fixed. Next, the transformer creates a query, a key, and a value vector for each word embedding, packs them together into matrices Q, K, V, and calculates a scaled dot-product attention (see Fig. 2.4) followed by non-linear

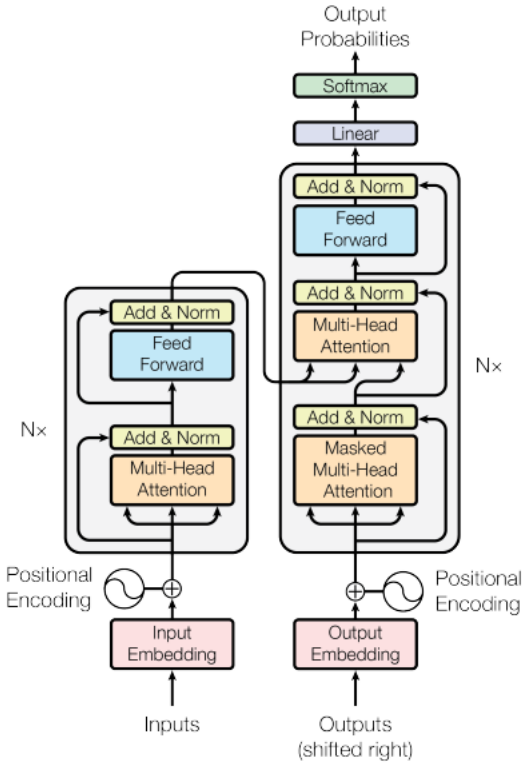


Figure 2.3: The Transformer - model architecture (taken from Vaswani et al. 2017).

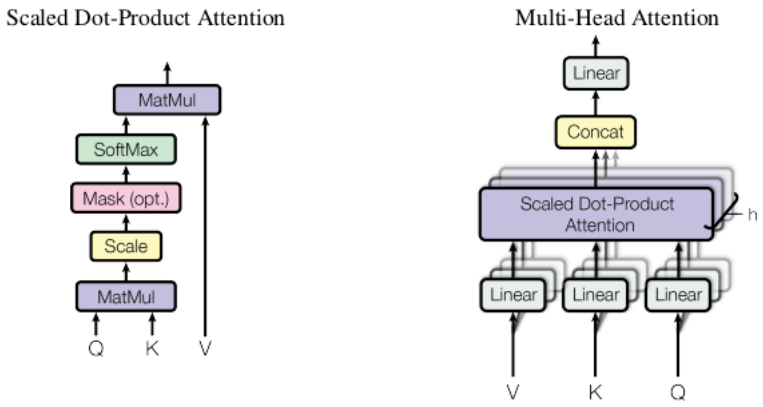


Figure 2.4: Scaled Dot-Product Attention. (left) Multi-Head Attention consists of several attention layers running in parallel. (right) (image taken from Vaswani et al. 2017)

transformation[2]:

$$Attention(Q, K, V) = softmax\Big(\frac{QK^T}{\sqrt{d_k}}\Big)V \tag{2.14}$$

The authors found it beneficial to linearly project the queries, keys and values h times and labeled it a multi-head attention (Fig. 2.4). This technique allows the model to jointly attend to information from different representation subspaces at different positions:

$$MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^O \tag{2.15}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2.16}$$

[2]In equation 2.14 d_k is the dimension of a key, and $\sqrt{d_k}$ has a role of a scaling factor. Vaswani et al. (2017) suspects that for large values of d_k, the dot products grow large in magnitude, pushing the softmax function into regions where it has extremely small gradients. The dot products are therefore scaled to counteract this effect.

Next, the outputs of the attention sub-layers are fed into a fully connected feed-forward network, which is applied to each position separately and identically:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{2.17}$$

The calculations are repeated for N layers with the only exception that instead of word embeddings the outputs of lower-level feed-forward neural networks are presented as the inputs to the layer above.

The decoder performs similar operations with a few modifications. The self-attention sub-layer in the decoder is modified to prevent positions from attending to subsequent positions. This is the reason why it is called a masked multi-head attention. The second modification of the self-attention sub-layer compared to the encoder is in the way how it creates queries, keys, and values. To connect the decoder with the encoder, and to help decoder focus on the appropriate word from the input, keys and values are created from the output of the top encoder, while queries are created from the output of masked multi-head attention.

Finally, the output of the top layer of the decoder is fed into the linear layer, followed by the softmax layer that outputs probabilities for each word from a vocabulary.

2.1.3 Abstractive summarizer

In this section, we describe the summarization model (Chen & Bansal, 2018) that we used in our work. The architecture of the model is relatively complex and belongs to hybrid approaches to text summarization that combine abstractive and extractive elements. On a high level, it consists of the extractive network (selects salient sentences), the abstractive network (rewrites or paraphrases them), and the reinforcement learning (RL) step that optimizes the model end-to-end. Both the extractor and abstractor networks are

learned independently, and during the RL step, the model updates only the extractor weights and leaves the abstractor as it is.

The goal of the extractor agent is to select salient sentences[3]. First, it uses a temporal convolutional model (Kim, 2014) to compute the representation for each sentence. To capture long-range dependencies between sentences, a bidirectional LSTM is applied to the convolutional output. This takes into account the context of all previous and future sentences in the same document. Finally, a Pointer Network (Vinyals et al., 2015) is added to extract sentences recurrently.

The task of the abstractor network is to compress selected sentences. It uses standard encoder-aligner-decoder (Bahdanau et al., 2015) with the copy mechanism from See et al. (2017). We have explained the components (LSTMs with the attention mechanism) of the abstractor in previous sections.

Both the extractor and abstractor are trained independently to minimize the cross-entropy loss. It is infeasible to train the whole model in an end-to-end fashion: the abstractor would get non-relevant sentences to compress from the extractor, and on the other hand the extractor would get noisy rewards from the abstractor.

After the training, the Advantage Actor-Critic (A2C reinforcement learner), a synchronous variant of A3C (Mnih et al., 2016), was used to optimize the model end-to-end maximizing the $ROUGE - L_{F_1}$ reward. Intuitively, if the extractor chooses a good sentence, a reward would be high, and thus the action is encouraged.

[3]The sentence selection task is formed as a classification task. Since the learning dataset contains only text-summary pairs and is not a classification dataset by itself, it has to be modified. The authors built a classification dataset for the extractor agent by finding the most similar text sentence for each ground-truth summary sentence using $R_{ROUGE-L}$ score.

2.2 Language models

A language model (LM) computes a probability for a sequence of words (Jozefowicz et al., 2016):

$$P(w_1, \ldots, w_n) = \prod_{i=1}^{n} P(w_i | w_1, \ldots, w_{i-1}) \tag{2.18}$$

For example, the sequence "the car is in the garage" has higher probability than the sequence "the garage is in the car". LMs are an integral part of various NLP tasks, e.g., speech recognition, text translation, text summarization, text generation, or prediction of missing words. Recently, it has been shown that generative pre-training of a language model, followed by fine-tuning on each specific task, improves state of the art on 9 out of 12 superGLUE NLP tasks (Radford et al., 2018). A good LM usually captures a grammar and other useful information from the text. It is not necessary to build a LM on a word level, it can also be built on a character or sub-word level.

An extrinsic and intrinsic evaluations of LMs exist (Jurafsky & Martin, 2014). The extrinsic evaluation measures the improvement of a given downstream application with a specific language model in comparison to a baseline language model. Intrinsic evaluation measures the performance of a LM independent of its use. In practice, the perplexity measure is used rather than raw word probabilities due to numerical stability. In information theory, perplexity is a measurement of how well a probability distribution or a probability model predicts a sample. The perplexity is sometimes used as a measure of prediction problem difficulty and is calculated as:

$$PP(W) = \sqrt[n]{\prod_{i=1}^{N} \frac{1}{P(w_i | w_1, \ldots, w_{i-1})}} \tag{2.19}$$

Neural approaches are currently state of the art for a language modeling task, especially the transformer based neural approaches (Dai et al., 2019; Radford et al., 2019).

2.3 Word embeddings

Neural prediction models can not process raw texts so the text elements have to be converted into a numerical representation. In the beginning of NLP, words were categorically encoded as a one-hot numeric arrays. This sparse representation is restrictive in its use and deals with raw frequencies or their transformations. For example, we can calculate the term-frequency–inverse-document-frequency weights that reflects how important a word is in a document.

Dense representations of words, called word embeddings, are real-valued vector representation of words, capturing both semantic and syntactic properties obtained from unlabeled large corpora (Wang et al., 2019). Many methods were proposed to calculate word embeddings: neural network language model (Bengio et al., 2003), neural continuous-bag-of-words and skip-gram models (Mikolov, Chen, et al., 2013), matrix factorization methods such as latent semantic analysis (Landauer et al., 1998), weighted least-squares model (Pennington et al., 2014), and recently deep contextual models such as BERT (Devlin et al., 2019), and ELMo (Peters et al., 2018). One difference between the model is how much information they use (local window-based information or global statistical information), and what the inputs of a model are, e.g., words, characters, sub-word information, n-grams, documents, etc.

Many intrinsic benchmarks for the quality of embeddings exist (Wang et al., 2019). The word similarity task correlates a cosine similarity of two word vectors and human perceived semantic similarity. Mikolov, Chen, et al. (2013) designed a comprehensive dataset that tests the quality of embeddings with semantic and syntactic questions (word analogy tasks), e.g., opposites, superlatives, plural nouns, currencies, gender, capital cities, etc. Other examples of benchmarks are concept categorization where the goal is to split words into proper categories, and outlier detection where the goal is to find words that do not belong to a given group of words.

The need to represent meaning in multilingual contexts and transfer knowledge to low-resource languages is on the rise (Ruder et al., 2019).

Cross-lingual embeddings are word vectors of two or more multiple languages aligned in the same vector space. Many methods to achieve such an alignment exist, supervised, semi-supervised, and unsupervised, and they usually presuppose independently trained monolingual word embeddings. One of the first proposed methods was minimizing the mean squared error (Mikolov, Le, & Sutskever, 2013) of the alignment:

$$\min_{W} \sum_{i=1}^{n} ||Wx_i - z_i||^2 \tag{2.20}$$

where W is a translation matrix, x_i is a source word and z_i is a target word. The source-target pairs must be provided by a dictionary, which makes the method supervised. Conneau et al. (2017) presented an unsupervised cross-lingual mapping with domain-adversarial approach.

Chapter 3

Datasets

In this chapter, we describe the process of creating two datasets for our needs, one for summarization task and the other for language modeling. Both datasets were extracted from the Gigafida 2.0 (2019) corpus of written standard Slovene, consisting of newspapers, magazines, and web texts. Less than 10% are fiction and non-fiction publications. The texts have been processed with the aim of creating a corpus that represents a sample of modern standard Slovene and can be used for developing language technologies. It contains 38,310 texts and more than 1.1 billion words.

3.1 STA summaries dataset

A training dataset of text and summary pairs was produced by taking the first paragraph of STA (Slovenian press agency) news web texts as a summary and the rest of it as a text. Since the Gigafida corpus from which we extracted STA news is sentence segmented but not paragraph segmented, we designed a heuristic to produce the learning dataset.

From the STA web site (STA, 2019), we downloaded 25 randomly chosen first paragraphs and calculated the average number of characters per paragraph. We appended one generic sentence, which represented the first sentence of a text, to each original paragraph. After that, we extracted the

Total news	Correct length	Underestimated	Overestimated
25	20	2	3

Table 3.1: Results of our heuristic to produce text-summary pairs. We achieved 80% accuracy for 25 randomly chosen first paragraphs. Two generated summaries were too short, which means that not all of the relevant sentences were extracted, and three of them were too long, which means that they contained the additional generic sentence.

sentences of a text one by one, approximating the calculated average length of a summary. The extraction procedure was terminated when the extracted sentence length deviated more from the calculated average than it was in the previous step. This heuristic produced the results presented in Table 3.1.

We collected 284,000 training samples but took only texts between 1000 and 3000 characters. Some texts were of no interest for us because they were comprised of weather reports, lists of events around the world, etc., and some of them were just too long. A total of 127,563 samples remained. They were split into train, test and validation sets. Both the test and validation sets contain 5000 news and the training set contains the remaining 117,563 samples.

3.2 The language model dataset

We built the dataset to train a language model to be used to fix errors due to cross-lingual transfer. We assumed that cross-lingual word mapping alone cannot produce correct texts in the target language because the grammar of a decoder remains in a source language. In Chapter 5, we present various ways how we used language models.

Our task was to extract sentences and preprocess them on a character level. The Gigafida corpus is already tokenized and sentence segmented. All punctuations, special characters, and numbers were preserved, but alphabetical characters were lower-cased. A total of 59,861,870 sentences were

extracted. The average sentence length is 242 characters with the standard deviation of 173. 95% of samples fall between 26 and 622 characters. The sentences were split into the train, test and validation set with ratios 90:5:5.

Chapter 4

Evaluation metrics

This chapter covers ROUGE and BERTScore metrics we used to evaluate candidate summaries, and precision @k metric we used to assess the quality of mapped word embeddings.

4.1 ROUGE

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a metric most commonly used for the evaluation of automatically generated text summaries. It measures the quality of a summary by the number of overlapping units (n-grams, sequences of texts, etc.) between summaries created by humans and summaries created by summarization systems (Lin, 2004). ROUGE is not a single metric but a family of metrics. The most commonly used are ROUGE-N and ROUGE-L. The first measures the overlapping of n-grams (typically unigrams and bigrams), while the second measures the longest common subsequence found in both summaries.

ROUGE is recall-oriented, contrary to precision-oriented BLEU. The original formula expresses the ratio between overlapping units $\sum matching\_units$ and the number of units in the reference summary $\sum reference\_units$

$$R_{ROUGE} = \frac{\sum matching\_units}{\sum reference\_units}, \tag{4.1}$$

usually, the F_1 version is reported. To calculate the F_1 variant, we need to calculate the precision which is the ratio between overlapping units $\sum matching\_units$ and the number of units in the system summary $\sum system\_units$

$$P_{ROUGE} = \frac{\sum matching\_units}{\sum system\_units}, \tag{4.2}$$

and then calculate the harmonic mean of precision and recall to get the F_1 score:

$$F_{1_{ROUGE}} = 2 \cdot \frac{R_{ROUGE} \cdot P_{ROUGE}}{R_{ROUGE} + P_{ROUGE}}. \tag{4.3}$$

The ROUGE metric can be misleading and inaccurate in many cases. For example, the two sequences

1. The black cat is on the mat.
2. The mat is on the black cat.

will have a perfect ROUGE-1 score although their meanings differ. On the other hand, sentences with similar meanings will have a zero score like in the following example:

1. It is cold.
2. Chilly weather persists.

4.2 BERTScore

BERTScore (T. Zhang et al., 2019) builds on BERT, which is a standard multi-layer bidirectional Transformer encoder (Devlin et al., 2019). BERT is comprised of the transformer components already described in Section 2.1.2. The model was pre-trained with tasks of masked language model and the next sentence prediction. It can be fine-tuned for many NLP tasks.

To get a BERTScore, we need to calculate the token representations and similarity measures between tokens of two texts. We use a pre-trained BERT

model[1] to generate the contextual token representations of the words in the candidate x and reference $\hat{x}$ sentences. In the next step, we calculate pairwise cosine similarity between the words and use greedy matching to maximize the similarity scores of recall, precision, and F_1:

$$R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} x_i^T \hat{x}_j, \tag{4.4}$$

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} x_i^T \hat{x}_j, \tag{4.5}$$

$$F_{BERT} = 2 \cdot \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}}. \tag{4.6}$$

The second case above, for example, where ROUGE-1 score is zero, the F_{BERT} scores 0.8491.

4.3 Precision @k

Here we explain only the @k part of the metric used to assess the quality of cross-lingual word embeddings, since the concept of precision has been already explained in 4.1. To calculate the metric, we need a polysemy dictionary, mapped embeddings, and similarity scores (nearest neighbours) for each word vector in a source language with each word vector in a target language. After that, we can measure how many times the correct translation of a given word was retrieved for top k neighbours and report the precision.

[1] In our work, we used bert-base-multilingual-cased_L9_no-idf_version=0.2.2

Chapter 5

Architecture and training

In this chapter, we first outline our solution to the problem of cross-lingual summarization. We provide detailed descriptions of word embeddings, the pre-trained English summarization model, our Slovene summarization models, the transformer language model, and possible alternatives we considered during the creation of the final solution scheme.

We created a Slovene dataset of summaries described in Section 3.1, and trained a Slovene language model on a very large Gigafida corpus described in Section 3.2. We used a pretrained English summarization model and extracted the English word embeddings from it. We used pretrained Slovene fastText word embeddings (Grave et al., 2018). Using the cross-lingual mapping technique described in Section 2.3, we mapped Slovene word embeddings into the English word vector space. Our cross-lingual models were trained with an increasingly large fractions of the created Slovene STA dataset. From models we gathered a large set of generated hypotheses with internal scores for each candidate sentence. We evaluated the hypotheses with an independently trained language model and two extra metrics. The best model was produced by a greedy search through available metrics. The solution scheme is presented in Fig. 5.1.

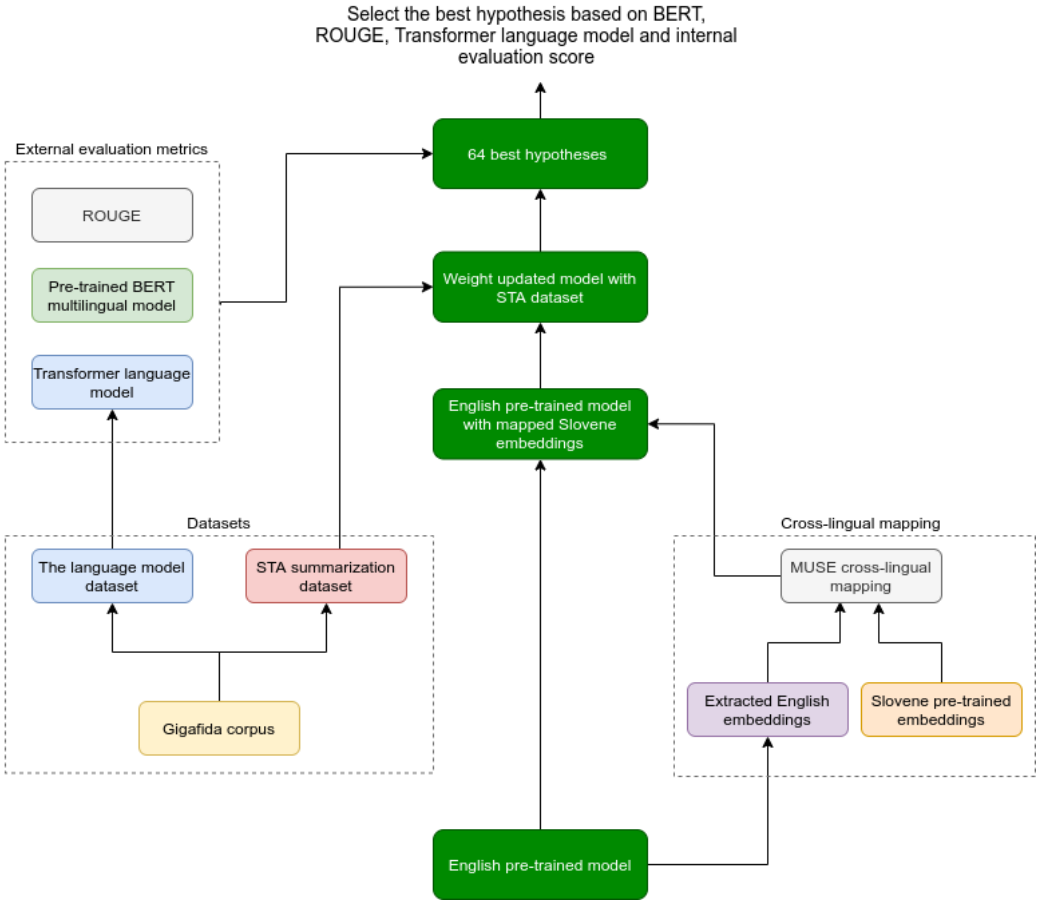


Figure 5.1: Illustration of how data is processed.

5.1 Word embeddings

We used pretrained fastText embeddings (Grave et al., 2018). The embeddings were trained on a mixture of Wikipedia and Common Crawl data. The model used is an extension of the word2vec CBOW model with position weights and subword information. We used the MUSE library (Conneau et al., 2017) for cross-lingual mapping. We described the mapping technique (Xing et al., 2015) in Section 2.3. The train dictionary of size 5000 and the test dictionary of size 1500 are part of the library and were created using an internal translation tool.

5.2 Summarization models

We used the pretrained summarization model proposed by Chen and Bansal (2018). We also considered two alternatives, i.e. Gehrmann et al. (2018) and See et al. (2017). The first one uses complex dynamic embedding layers inappropriate for our task due to alignment complexity: a combination of static and contextual embeddings is used; although there are also methods proposed for the alignment of contextual embeddings (Schuster et al., 2019). The second one was not publicly released. Chen and Bansal (2018) approach trained custom word2vec embeddings and was chosen because it allows simple cross-lingual mapping, uses dynamic extractive-abstractive approach, and offers the same level of performance as other state of the art summarization models on English.

The model is trained on the CNN/Daily Mail dataset, a standard for testing summarization systems. The dataset contains online news articles (781 tokens on average) paired with multi-sentence summaries (3.75 sentences or 56 tokens on average). The processed version contains 287,226 training pairs, 13,368 validation pairs, and 11,490 test pairs (Nallapati et al., 2016).

The components of the model were described in Chapter 2, here we present only the details. Skip-gram training algorithm was used to calculate word embeddings with window size of 5. The dimension of the embeddings

was 128, and the size of the vocabulary was 30,000. Word vectors are fed through 1-D single-layer convolution filters with various window sizes (3, 4, 5), followed by ReLU non-linear activation and max-over-time pooling. The convolutional representation r_j for the jth sentence is obtained by concatenating the outputs from the activations of all filter window sizes. A one layer bi-LSTM with 256 units is applied on the convolutional output and one layer LSTM (unidirectional) with 256 units is used to select sentences.

The abstractor is a sequence-to-sequence model with attention (Bahdanau et al., 2015; Luong et al., 2015) and copy mechanisms (See et al., 2017) to predict over the extended vocabulary of words in the input document. To calculate the attention scores, the bilinear multiplicative attention function is used. The source and target embeddings are shared. Both encoder and decoder networks are comprised of one layer with 256 LSTM units.

Advantage Actor-Critic, a synchronous variant of A3C (Mnih et al., 2016) is used to optimize the whole model with the discount factor $\gamma = 0.95$.

Both the extractor and abstractor learning rate was 0.001, and for reinforcement learning optimization step was 0.0001. No regularization was used except early stopping. Adam optimizer was used, with the learning rate halved when validation loss stop decreasing.

Training took 4.15 hours for the abstractor, 1.5 hours for the extractor, and 15,56 hours for the RL optimization on Nvidia Tesla K40 GPUs.

5.3 Our models

Since we experimented with transfer learning, we did not change the architecture described in the previous section. Our models differ only in proportions of the dataset that were used for weight updates. We also replaced the English dictionary with the Slovene dictionary which was built from 30,000 most common words in the STA dataset.

We compare three models: zero-shot, weight-update, and Slovene model. The details are given below and summarized in 5.1.

Model	Slovene data size	Details
MENG	0%	no weight updates, zero-shot transfer learning
M1	1%	trained extractor, weight update abstractor
M10	10%	trained extractor, weight update abstractor
M100	100%	trained extractor, weight update abstractor
MSLO	100%	trained extractor, trained abstractor

Table 5.1: Summary of produced summarization models.

Zero-shot model is a baseline model. No Slovene data was presented to it. In the pretrained model, only the English embeddings were swapped with the aligned Slovene embeddings.

Weight update models We built three datasets composed of 1%, 10%, and 100% of the STA dataset. We trained the extractor of each model because only the reinforcement learning optimized extractor was provided by the authors. Simultaneously, we updated the weights of the pretrained abstractor and, finally, the models were optimized with RL.

Slovene model All components of this model were trained on the complete STA dataset from scratch.

5.4 Language model training

We trained the language model with the aim of output sentence evaluation. We trained the Slovene language model on the character-level rather than word-level since Bojanowski et al. (2017) found that the models for morphological rich languages are improved by using character-level information.

Many current state-of-the-art language models (Baevski & Auli, 2019; Dai et al., 2019), trained on a dataset similar to ours (Chelba et al., 2014), use variants of the transformer architecture. We used a transformer decoder as implemented in the Tensor2tensor library (Vaswani et al., 2018), Adam

optimizer, 8 attention heads, 6 hidden layers, and position-wise feed-forward networks with one hidden layer of size 2048 and ReLU activation function. These are standard hyperparameters for training on a single GPU. We increased the maximum size of the input from 256 to 512, based on the statistics of sentence lengths in the learning corpus. We calculated 95th percentile to be 528 and 512 chosen. Shorter sentences are padded and longer are cut off. The dictionary contains 581 characters. The total number of learning parameters was 19,035,136.

The language model was trained for 1,500,000 steps in two parts (900,000 and additional 600,000) due to limited resources. One step means one weight update with batch size of 2048. This means 60 epochs for the first part in 40 epochs for the second part. The training set was split into approximately 30 million and 23 million sentences. The model was evaluated on a test set with size 10k. Training took approximately 4 days on Nvidia Titan X 12GB GPU.

5.5 Postprocessing the summaries

Here we present how we postprocessed hypotheses produced by the models, and chose the best among them. We took a rule-based approach to eliminate repeating n-grams and additional external metrics to select the best hypothesis.

5.5.1 Rule-based correction

Sometimes models struggle with repeating n-grams:

> ”kanadska smučarka smučarka 2016 2016 v kanadi , kanadi , china alpska alpska alpska 2016 za kanadi , kar je to dobra vzpodbuda za naprej za naprej . za”

This problem was solved by eliminating repeating n-grams with rule-based approach. After the eliminating function is applied, we get:

"kanadska smučarka 2016 v kanadi , china alpska 2016 za kanadi , kar je to dobra vzpodbuda za naprej . za"

5.5.2 Additional metrics

Besides the internal evaluation of hypotheses using the loss function, we decided to evaluate them externally using ROUGE, BERTScore (T. Zhang et al., 2019), and independently trained language model scores. We calculated ROUGE and BERTScore scores by comparing generated hypotheses with the extracted sentence.

We produced the final model by combining two metrics. We assumed that the performance may be improved by first checking the content of generated hypotheses (BERTScore or ROUGE), and then the formulation of them (by independently trained language model or the internal evaluation scores of the decoder). The procedure is simple: the model first generates a set of n hypotheses. From these n hypotheses, it selects m best content-wise. Finally, the best of m is selected readability-wise.

5.6 Alternative approaches

Lastly, we describe two alternative approaches to improve the hypotheses produced by the models.

Sentence correction by rules

We tried parsing a sentence and use the information in various ways. For example, we tried to identify the missing parts of speech in the summary and add them from the source sentence. We also tried to define the sentence structures of the two sentences, compare them and then remove or add words. With that, we fell into the endless correction task, typical of such linguistic approaches. Moreover, such corrections are outside the central theme of the work. We are convinced that the solution must rely on neural approaches

Number	Hypotheses
0	the black cat is on the green
1	the black cat is on the mat
3	the black cat is on green mat
4	the black cat is the green mat
5	the black cat on the green mat
6	the black is on the green mat
7	the cat is on the green mat
8	black cat is on the green mat

Table 5.2: Hypotheses for the sentence "the black cat is on the green mat". Only the best n hypotheses are retained.

with the exception of simple corrections as described above. Nevertheless, this may be an interesting approach for shortening individual sentences.

Missing words prediction

We found the missing words prediction a much more promising approach. The idea is similar to beam search. We propose making one-word shorter hypotheses of a sentence and evaluating them with a language model as demonstrated in Table 5.2. Some of the best candidates are retained, and re-evaluated; the process continues until the minimum sentence length is reached.

The first results show that we can shorten the sentence in this way, but the input must be a grammatically correct sentence, otherwise the model gets confused. This method could be improved by part of speech tokenization rather than single words. We can also invert the task and try to construct the shortened variant of a sentence from the extracted sentence by adding words.

Chapter 6

Results

In this chapter, we first provide the results for the quality of the used components: cross-lingual mappings, and the transformer language model, followed by the results of summarization. In addition to automatic evaluation, we present human evaluation of the generated summaries. We discuss a few system-generated summaries, and compare our results with related works.

6.1 Cross-lingual mapping

We mapped the pre-trained Slovene word embeddings to the English vector space of the pre-trained summarization model with the Procrustes alignment method described in Section 2.3. We used a word translation task to calculate the quality of the mapped embeddings. The task evaluates how successful is the translation of a source word. We evaluated the embeddings on 1,500 samples large Slovene-English dictionary that deals well with the polysemy of words (Conneau et al., 2017). In Tables 6.1 and 6.2, a precision @k is reported, where k is the size of the neighborhood. The evaluation metrics are nearest neighbors (NN) and cross-domain similarity local scaling (CSLS). CSLS is an improved variant of NN that mitigates the problem of hubs (vectors with high probability are nearest neighbors of many points, while some points are not nearest neighbors of any point).

Eval	P@1	P@5	P@10
NN	27,03	45,77	53,99
CSLS	29,47	48,04	55,78

Table 6.1: Abstractor mapping results from English to Slovene.

Eval	P@1	P@5	P@10
NN	19,87	37,54	45,27
CSLS	21,09	40,31	46,58

Table 6.2: ExtractorRL mapping results from English to Slovene.

Using the same method as ours, Conneau et al. (2017) reported scores above 80 on P@1 from English to Spanish or French and above 60 on P@5 and P@10 from English to Italian. The quality of our ExtractorRL mapping results are similar to the Italian to English mappings (Mikolov, Le, & Sutskever, 2013). Our Abstractors results can compare to the English to Italian ones reported by the same authors.

We could not identify the reason why ExtractorRL embeddings quality is significantly lower than Abstractors. We also tried to align the embeddings without cross-lingual supervision using a domain-adversarial setting (Conneau et al., 2017), but we were not successful and did not find a reason why.

6.2 Language model

We report the performance of our language model in Table 6.3. Our results are not comparable to the best English language models. The state-of-the-art model GPT-2 scores 0.93 on the enwik8 dataset and 0.98 on the text8 dataset (Radford et al., 2019). GPT-2 has 1542M parameters and is thus approximately 80 times larger than our model. The model, similar in size to ours, scores 1.34 on the text8 dataset (Ha et al., 2016). The best scores vary

Step	bits/char
900k	1,835
1,500k	1,787

Table 6.3: Transformer language model results on test data after training for 900k steps and 1,500k steps. The loss function used is a character-level cross-entropy (bits/char). Lower scores mean better language model

with respect to dataset from 0.93 to 1.15. Due to morphological complexity of Slovene, the comparison is not entirely fair, but currently, no other publicly available Slovene language model is available for comparison.

Qualitative analysis

Considering the hypotheses in Table 6.4, it is evident that our Slovene LM generates well-formulated and meaningful hypotheses. In hypothesis 7, the model did not correctly finish the sentence; the dot should not be present, since each sentence has to be completed with a special end of sentence token. Hypotheses 11-16 contain the quotation marks which should not be there because they are not present in the given sentence. However, we may interpret it as an attempt to insert a quote.

Table 6.5 shows more examples, and the best hypothesis for each of them. All examples are grammatically correct with the commas in the right place. All words are well-formulated. Meaning can be convoluted, amusing, or wrong. It is convoluted how American oil can lose profits (example 1), amusing that police officers were pleasantly prepared on Saturday and Sunday (example 4), and wrong (example 6) that professions can make decisions (especially decisions to make decisions). The model can also recognize well-formulated sentences as inputs, and not output anything (Example 2). It is noticeable that the outputs are mostly of a political and financial nature since the corpus used to train LM consists mainly of daily news.

Number	Hypothesis	Score
1	želim	1.348512
2	morem privoščiti	1.579529
3	moremo privoščiti	1.682322
4	morem predstavljati	1.762461
5	predstavljam	1.832580
6	moremo predstavljati	1.891039
7	morem privoščiti .	1.929132
8	moreš privoščiti	1.936325
9	moreš predstavljati	1.966972
10	smemo privoščiti	1.988507
11	morem privoščiti , ” je dejal	2.398079
12	moremo privoščiti , ” je dejal	2.446752
13	morem predstavljati , ” je dejal	2.491814
14	moremo predstavljati , ” je dejal	2.573889
15	moremo privoščiti , ” je povedal	2.650967
16	morem predstavljati , ” je povedal	2.701185

Table 6.4: Examples of sentence completion task. We presented the sentence ”tega si nikakor ne” to the language model which generated 16 possible hypotheses. The score column is the result of the negative log likelihood loss function.

Num	Input	Output
1	ameriška nafta	je namreč v zadnjih desetih letih izgubila približno tri milijone evrov čistega dobička
2	za vsaj en večer se je uresničila	
3	nekaterim je	pomagala , drugim pa je bilo treba povedati , da je bila njena prijateljica
4	policisti so bili v soboto	in nedeljo popoldne prijetno pripravljeni
5	nove članice evropske unije	so se odločile , da bodo predstavniki evropskega parlamenta nadaljevali prihodnje leto
6	poklici , ki jih ne	enakopravno predstavljajo , so se odločili , da se bodo odločili , ali bodo predstavniki državnega zbora
7	to pomeni ,	da je treba v primerjavi z lanskim letom povečati število zaposlenih
8	tudi s	lovenski predsednik republike borut pahor je prepričan , da bo predsednik uprave za zdravstveno zavarovanje in zdravstveno zavarovanje zavarovalnice triglav
9	slovesnosti v münchnu se je udeležilo okoli	1000 predstavnikov slovenskega predsedniškega kandidata

Table 6.5: Examples of sentence completion task. We built test cases by randomly taking 1 to 10 initial words or characters from randomly selected sentences in the test set. We qualitatively assessed a small sample and found that the model has to consider between 12 and 16 hypotheses to give the best possible completion of a given sentence.

6.3 Summarization models

In this section, we report the results of summarization models. We calculated the ROUGE scores using the pyrouge package (2020).

Table 6.6 shows summary statistics of the generated summaries. The English model MENG generates more than twice as many characters as the other models. It generates 4 sentences while the other models generate 2 to 3. Note that the number of extracted sentences is result of the learning. The average number of generated sentences per summary thus succesfully reflects the average number of summary sentences. M1 shows that as little as 1k additional examples are enough to update the number of extracted sentences. In our case, from 4 sentences present in CNN/Daily Mail dataset to 2-3 sentences in STA dataset. These observations can be seen in Appendix A.

Table 6.7 shows the results of the baseline models explained in Section 5.3. Tables 6.8 and 6.9 present optimization of the best model by tuning one or two parameters, respectively. The first reason why MENG model scored higher on ROUGE than M1 and M10 is that it extracts more sentences, generates longer summary sentences, and repeats the sentences. The second reason is that it has a fully trained extractor agent which shows that model transfer was successful when considering which sentences to extract. Analysing the results of MENG, we noticed that the model sometimes cannot finish a sentence properly, e.g., it generates good content, but can not stop and just continues generating words. We speculate that the problem lies in special tokens (start of sentence, end of sentence, etc.) that capture the grammar of source language. These special tokens may be a hidden problem in the cross-lingual seq2seq research field.

It is tough to make definitive conclusions regarding readability of the models. M1 does not show any significant readability improvement over MENG, while M10 shows some improvement (see Appendix A). MENG often generates long sentences with redundant and rare words, and inserts punctuations in inappropriate places. On the contrary, M1 generates too short sentences

Model	Sentences	Characters
MENG	3,99	500,61
M1	2,81	218,48
M10	1,95	204,59
M100	2,79	297,67
MSLO	2,58	270,79
Reference	2,10	302,02

Table 6.6: Average number of extracted sentences and average number of characters per summary. The last row represents the ground truth.

Model	ROUGE-1	ROUGE-2	ROUGE-L
MENG	18,91	3,74	16,27
M1	12,94	1,96	11,61
M10	15,71	3,71	13,87
M100	**21,67**	**6,81**	**19,16**
MSLO	21,07	6,62	18,64

Table 6.7: Results of non-optimized models.

Parameter 1	ROUGE-1	ROUGE-2	ROUGE-L
Baseline M100	21,67	6,81	19,16
Transformer	22,53	6,83	19,61
BERTScore	24,87	**7,41**	21,36
ROUGE-L	**24,88**	7,38	**21,47**

Table 6.8: Optimization of the best model by one parameter.

Parameter 1	Parameter 2	ROUGE-1	ROUGE-2	ROUGE-L
ROUGE-L	BERTScore	24,97	7,43	21,50

Table 6.9: Optimization of the best model by two parameters.

and summaries with a lot of missing words. M10 shows an improvement in sentence selection over M1, and readability over both M1 and MENG. Still, most of the sentence are not well-formulated, but we can say that the meaning is present in almost all of them.

M100 and MSLO are the best models. Note that M100 is a cross-lingual model and MSLO was trained from scratch. It is infeasible to conclude which model is better in terms of readability (see Appendix A). M100 shows better ROUGE scores in Table 6.7. ROUGE-1 is improved for 0,60; ROUGE-2 for 0,19; and ROUGE-L for 0,52. This suggests that our cross-lingual approach yields better results.

Table 6.8 shows optimization of the best model (see Section 5.5.2 which explains the details of the optimization procedure and the meaning of the parameters). Using transformer based LM, the model improves internal evaluation score by 0,86 point on ROUGE-1 and 0,45 point on ROUGE-L. Both BERTScore and ROUGE-L are indiscernible in performance.

Initially, our idea was to use two parameters to select the best hypothesis, one for content selection and one for readability. The results in Table 6.9 show that this is not the case. Using two best parameters, both of them belong to the content selection group. We are aware that these results may be biased since the final evaluation ROUGE metrics are content based. Qualitative analysis confirmed that using two parameters produced more readable summaries with lower content accuracy.

6.4 Examples of generated summaries

In this section, we present examples of summaries of non-optimized models from Table 6.7 and the example summary of our best optimized model from

Table 6.9.

6.4.1 Non-optimized models

Text

hrvaška tiskovna agencija hina je poročala o srečanju predsednikov držav pobude brdo brioni na brdu pri kranju . hina je povzela pisanje sta , da so sprejeli deklaracijo , v kateri so poudarili , da mora zahodni balkan ostati v središču interesa bruslja in članic eu in da je treba odprta vprašanja rešiti s političnim dialogom . srbska tiskovna agencija tanjug je pisala , da morajo biti vrata eu odprta za nove članice . o srečanju pa je poročala tudi avstrijska tiskovna agencija apa . tanjug je pisal , da sta se ob robu srečanja na brdu sestala tudi slovenski zunanji minister karl erjavec in srbski zunanji minister ivica dačić . izrazila sta zadovoljstvo nad večinoma dobrimi odnosi , govorila pa sta tudi o naslednji skupni seji obeh vlad , ki naj bi se odvijala jeseni v sloveniji . tanjug je poročal , da sta slovenski in hrvaški zunanji minister karl erjavec in davor ivo stier na srečanju zunanjih ministrov držav procesa brdo brioni povedala , da si bosta slovenija in hrvaška s skupnimi močmi prizadevali , da ostane na agendi eu širitev na zahodni balkan . hina je povzela poročanje srbskih medijev , da je srbski predsednik aleksandar vučić na srečanju brdo brioni načel temo gospodarskega povezovanja v regiji . tanjug je pisal , da je bila zanj to prva mednarodna dejavnost po prevzemu položaja v sredo . hina je poročala , da sicer na srečanju na brdu o arbitraži o meji med slovenijo in hrvaško niso govorili , a je bila tema novinarskih vprašanj . slovenski predsednik borut pahor je rekel , da bi neizvajanje arbitražne sodbe v mejnem sporu med slovenijo in hrvaško lahko povzročilo spodkopavanje moralne avtoritete ustanoviteljev procesa brdo brioni . o tem je poročal tudi tanjug . hina je pisala , da je slovenski zunanji minister karl erjavec poudaril , da je arbitražni sporazum veljavna mednarodna pogodba in da morajo takšne pogodbe spoštovati vse članice eu . hina je poročala , da je hrvaška storila vse , kar je bilo potrebno , za izstop iz arbitraže s slovenijo

. o tem je poročal tudi tanjug . hina je poročala , da slovenski mediji menijo , da je nemčija zavzela nevtralni položaj v arbitrarnem sporu med slovenijo in hrvaško .

MENG

hrvaška tiskovna agencija hina je poročala o srečanju predsednikov držav pobude brdo brioni na brioni vič brioni brdu pri
srbska tiskovna agencija tanjug je pisala , da morajo biti vrata eu odprta za nove . članice .
hina je povzela pisanje sta , da zahodni balkan ostati v interesa bruslja in članic eu in da mora odprta vprašanja rešiti s političnim
poročala tudi je poročala tudi avstrijska tiskovna apa . avstrijska tiskovna agencija apa
ROUGE: 50,00

M1

tanjug je danes poročal , poroča francoska tiskovna agencija afp
tuje tiskovne agencije
ROUGE: 8,45

M10

tuje tiskovne agencije so med drugim poročale o srečanju predsednikov držav pobude
poročale so tudi , da je zahodni balkan ostati v eu
poročale so tudi o meji med slovenijo in hrvaško
poročale so tudi o tem , da sta slovenski zunanji minister karl erjavec in davor ivo stier na srečanju zunanjih ministrov eu
ROUGE: 39,29

M100

tuje tiskovne agencije so med drugim poročale o srečanju predsednikov držav pobude brdo brioni na brdu pri kranju
pisale so tudi o tem , da mora zahodni balkan ostati v središču interesa bruslja in članic eu
ROUGE: 68,09

MSLO

tuje tiskovne agencije so med drugim poročale o srečanju predsednikov držav pobude brdo brioni
pisale so tudi o tem , da mora zahodni balkan ostati s političnim dialogom
ROUGE: 55,81

Golden reference

tuje tiskovne agencije so danes med drugim poročale o srečanju predsednikov držav pobude brdo brioni na brdu pri kranju . hina je povzela pisanje sta , da so sprejeli deklaracijo , v kateri so poudarili , da mora zahodni balkan ostati v središču interesa bruslja in članic eu in da je treba odprta vprašanja rešiti s političnim dialogom .

6.4.2 Examples from the best model

In this section, we analyze a few examples of our best model. This model summarizes texts with a variate quality. We believe that the most important explanatory factor for the differences in quality is the topic of a text. The model generates good summaries for texts with political and financial content, which represent the majority of our dataset.

The first example demonstrates one of the best produced summaries. It is short, contains essential information, contains well-formulated sentences, and exhibits certain level of abstraction. It replaces the phrase "Hrvaška tiskovna agencija hina je pisala o tem" with "tuje tiskovne agencije so poročale o tem"

and cuts off additional information that starts with "ob napovedi avstrije ... ". The second sentence exhibits similar qualities, but "predsednik evropskega sveta donald tusk" is omitted for no apparent reason. The second example shows how the model can be misleading. The play is not in Ljubljana but in Maribor. The model speculates that the play will start at 8 p.m, although correctly uses "drevi" as "in the evening". The content is repeated. However, the third sentence is impressing. The third example shows that the model correctly identifies who defeated who, but completely misrepresents the numbers, which was one of the most frequent errors observed.

Example 1

hrvaška tiskovna agencija hina je pisala o tem , da je slovenska vlada izrazila negativno mnenje o avstrijskem nadzoru na meji s slovenijo ob napovedi avstrije , da bo za še eno polletno obdobje podaljšala nadzor na notranji schengenski meji s slovenijo . o tem je pisala tudi avstrijska tiskovna agencija apa . hina je poročala tudi o tem , da se bo slovenski premier marjan šarec na jesenskem uradnem obisku v bruslju srečal s predsednikom evropskega sveta donaldom tuskom in predsednikom evropske komisije jean - claudom junckerjem . slednji je v zadnjem času v ljubljani deležen številnih kritik zaradi domnevne pristranskosti v arbitražnem sporu med slovenijo in hrvaško . hina je pisala še o tem , da slovenski zunanji minister miro cerar , ki je trenutno v washingtonu , pričakuje izboljšanje odnosov med slovenijo in zda . američane namerava bolje seznaniti z arbitražnim sporom med slovenijo in hrvaško , saj s tem problemom po njegovem mnenju niso dovolj seznanjeni . srbska tiskovna agencija tanjug je poročala , da bosta slovenska policijska sindikata policijski sindikat slovenije in sindikat policistov slovenije s ponedeljkom znova začela izvajati stavkovne aktivnosti , ki so zamrznile v marcu . tanjug je pisal tudi , da je srbski predsednik aleksander vučić danes sprejel slovenskega veleposlanika v srbiji vladimirja gaspariča na poslovilni obisk . ob tej priložnosti je gasparič izrazil prepričanje , da je bilo načrtovanje obiska , o katerem sta se vučić in slovenski predsednik borut pahor pred ne-

davnim pogovarjala , dodatna spodbuda za dobro sodelovanje med državama .

Reference: tuje tiskovne agencije so med drugim pisale o tem , da je slovenska vlada izrazila negativno mnenje o avstrijskem nadzoru na meji s slovenijo ob napovedi avstrije , da bo podaljšala nadzor na notranji schengenski meji s slovenijo .poročale so tudi , da bosta slovenska policijska sindikata znova začela izvajati stavkovne aktivnosti .

Candidate: tuje tiskovne agencije so poročale o tem , da je slovenska vlada izrazila negativno mnenje o avstrijskem nadzoru na meji s slovenijo . poročale so tudi , da se bo slovenski premier marjan šarec na jesenskem uradnem obisku v bruslju srečal s predsednikom evropske komisije jean - claudom junckerjem .

ROUGE: 51,46

Example 2

veličastna igra senc in zvoka je v rokah animatork barbare jamšek in elene volpi . predstava , ki bo premierno prikazana v četrtek zvečer , je nastala po motivih slikanice dennisa haseleyeja o piratu , ki je skušal ujeti luno ter s pesmimi bine štampe žmavc . ” zgodba govori o pohlepnem piratu , ki želi ukrasti cel svet , na koncu pa sega še po luni , ” je na današnji novinarski konferenci povedal režiser tin grabnar . takšna zgodba je po njegovem mnenju odlično izhodišče za predstavo senčnega gledališča , kjer se materialni svet v obliki lutk in drugih rekvizitov postavi v razmerje z nematerialnim v obliki svetlobe in senc . predstava poteka na ladji z dvema jadroma , ki je postavljena v nedavno obnovljeno cerkev , in služi hkrati kot oder in tribuna za gledalce . gledalec je postavljen v središče dogajanja in ima po besedah avtorice likovne podobe darke erdelji občutek , da je na morju , ” omejen z materijo , a s hlepenjem po več ” . glavni jezik predstave so sence , ne besede , saj so besedilo slikanice močno oklestili , da bi , tako dramaturginja

katarina klančnik kocutar , dosegli večji učinek prikazanega nasprotja med materialnim in nematerialnim . ” prilaščati si vse materialno , hkrati pa si želeti še več , še nematerialno , ” je glavno vodilo zgodbe z liki , ki so že od nekdaj burili človeško domišljijo , na primer luna , morje , pirati . slednji niso le simbol pohlepa po materialnih stvareh , ampak , glede na zgodovino , tudi ljudi na robu družbe , preganjanih iz različnih razlogov . pri ustvarjanju atmosfere igra zaradi odsotnosti besedila veliko vlogo glasba , katere avtor je iztok drabik jug , ki je med drugim posegel tudi po električni kitari . igralki in animatorki sta barbara jamšek in elena volpi . zanju je bila velik izziv poleg spoznavanja igre senc in uporabe svetil še igra na vse strani , saj sta v predstavi obkroženi z občinstvom . ker to ni klasična predstava senčnega gledališča , kjer so animatorji skriti za paravanom , je velika teža na koreografiji in gibu . pri ustvarjanju predstave so imeli nekaj težav z akustiko , saj cerkvi , ki se sicer drži lutkovnega gledališča in je bila obnovljena lani z evropskimi sredstvi , manjka tehnična oprema . po besedah direktorice gledališča mojce redjko so iskali prav takšno , ambientalno predstavo , da bi lahko izkoristili danosti sakralnega objekta , hkrati pa preizkusili pogoje dela v njem . sicer pa še vedno čakajo na občinski razpis za vsebinsko zapolnitev prostora .

Reference: marec v lutkovno gledališče maribor prinaša premiero predstave pirat in luna v režiji tina grabnarja .senčno gledališče , ki se posveča razmerju med materialnim in nematerialnim , so postavili v minoritsko cerkev , pri čemer cerkvena ladja služi kot plovilo

Candidate: v predstava teatru ljubljana) bodo drevi ob 20. uri premierno uprizorili predstavo dennisa haseleyeja o piratu , ki je nastala po motivih slikanice haseleyeja
v veličastna teatru ljubljana (veličastna) bo drevi premiera plesne predstave senc in zvoka
zgodba govori o pohlepnem piratu

ROUGE: 9,30

Example 3

formis : desy vahen 14 , sešel 5 , bračko , pintarič 20 , stavbar , polanec , geratič , gajser 15 , petranović 2 , štern 2 . kema puconci : praprotnik flisar 6 , vidonja 12 , vinkovič , franko 8 , škraban 12 , banfi 2 , donša 4 , jerič , koler , frumen 9 , sakovič 17 , kadiš 2 . odbojkarice formisa so doživele nov poraz , keme puconci pa so se veselile nove zmage . rogožanke z osvojeno točko niso naredile koraka proti sredini lestvice , zato pa so se prekmurke z dvema točkama obdržale v njeni zgornji polovici . uvodni niz je mineval v izenačeni igri , v končnici pa so gostje prikazale bolj zrelo igro od gostiteljic in povedle z 1:0 . tudi v drugem nizu si nobena ekipa ni priigrala občutnejše prednosti , rogožanke so izboljšale igro v napadu in obrambi , po vodstvu 24:23 in rezultatu 25:25 pa so osvojile še dve točki in izenačile rezultat v nizih . tretji niz so spet osvojile prekmurke , ki so s pridom izrabile slabe začetne udarce gostiteljic , v končnici pa so bile dovolj zbrane , da niso dovolile zasuka . domača ekipa je četrti niz začela zelo poletno in ves čas vodila . pri rezultatu 16:14 je zagospodarila na igrišču , nizala točke kot po tekočem traku in izenačila rezultat na 2:2 . odločilni niz so veliko bolje začele gostje , ki so povedle s 5:1 in 8:5 , v nadaljevanju so gostiteljice vzpostavile ravnotežje na igrišču , pri rezultatu 9:9 pa so zaradi napak pri sprejemu in v napadu dovolile gostjam , da se odlepijo za tri točke in tudi zmagajo .

Reference: odbojkarice kema puconcev so v tekmi 7. kroga 1 .dol za ženske v hočah premagale domači formis s 3:2 (21 , - 25 , 21 , - 16 , 15) . * športna dvorana v hočah , gledalcev 130 , sodnika : valentar (ravne) in štumfelj (mežica) .

Candidate: odbojkarice keme puconci so v 3. krogu 1 dol za ženske v gosteh premagale formis z 1:0 (1:0 * športna dvorana , gledalcev 250 , sodnika : bračko (kranj , štern odbojkarice kema puconci so v 2. krogu 1

ROUGE: 40,00

Score	Accuracy	Readability
1	none	incomprehensible
2	little	poor
3	a lot of	acceptable
4	most of	good
5	all	flawless

Table 6.10: Evaluation scales for accuracy and readability of summaries.

Type	Accuracy	Readability
Reference	2,85 (1,24)	4,18 (0,96)
System	3,06 (1,18)	3,41 (0,94)

Table 6.11: Average and standard deviation of accuracy and readability of reference and system summaries.

6.5 Human evaluation

Because of the problems of automatic evaluation, we decided to use human judgment to evaluate generated summaries.

We modified a human evaluation method described in Zidarn (2019). Instead of one summary per text, we used both reference and candidate but in a random order for each text. In this way, the evaluation was slightly modified but still enables comparison. The task of referees was to assign accuracy and readability of a summary (see Table 6.10 for the scales). Accuracy represents the amount of overlap between the given facts and summarized information, and readability measures how comprehensive a summary is. In our study, 10 articles (2 summaries per text) were evaluated by 8 referees. Referees include 3 females and 5 males aged from 23 to 65, and with IV. to VII. level of education.

We report averages and standard deviations in Table 6.11. It is surprising that the accuracy of the reference summaries is lower than the accuracy of the system summaries. We identified three reasons that explain this result.

Type	Accuracy	Readability
Reference	0,07	-0,06
System	0,07	0,05

Table 6.12: Fleiss' Kappa assesses the reliability of agreement between raters.

First, the reference summaries often contain true facts and information that cannot be verified by the text alone. Unless misleading and speculative, the system summaries should always produce verifiable content. Second, the evaluation method does not directly measure the content quality of a summary. Following the instructions, participants may assign a high score for a summary that contains true but unimportant and irrelevant information. Third, our model is a hybrid model which selects and paraphrases sentences. We assume that participants can be more easily lured into thinking that there is a greater overlap of content between a text and a generated summary than it is between a text and a reference summary. As we anticipated, the readability score of the reference summaries is much higher than it is for the system summaries.

We provide the Fleiss' Kappa scores in Table 6.12. This metric measures the agreement between the raters. We observed a slight agreement between the raters on all occasions, except the poor agreement on the reference readability situation. The drawback of Fleiss' Kappa is that it cannot take into account the ordinal nature of our variables. For example, two raters are assigned the same Kappa score, whether they rate a sample with the scores 4 and 5, or with the scores 4 and 1. Intuitively, the first pair of scores should have a higher agreement than the second pair. In spite of that, we believe the metric is the most appropriate for our task.

6.6 Comparison with related research

We compare our summarization model to related models for English and Slovene languages. Table 6.13 shows the results reported by authors. In addition to standard ROUGE scores, we also provide BERTscore.

Our model is most similar to Zidarn (2019) who used a two layer LSTM with attention mechanism, copy mechanism, and beam search. We both used the STA dataset but with a different train, test, and validations splits. Our model scored higher on ROUGE-1 (1,20) but lower on ROUGE-2 (0,54) and ROUGE-L (2,45). Interestingly, BERTScores were identical. Given the variation (different subsets of the original data, different splits and problematic nature of automatic summary evaluation metrics) we can conclude that both models perform similarly. Considering human evaluation, both models produce acceptable readability scores. In terms of accuracy, it seems that our model generates more accurate content.

Slovene models can not compare to English in terms of performance. English models are usually trained either on the 4 million examples large Gigaword dataset, appropriate for headline generation, or the 290k CNN/Daily Mail dataset, which is similar to our STA set. The English model used in our experiments (Chen & Bansal, 2018) achieves scores that are almost twice as large as ours. Its results are less misleading and usually represents facts and information accurately. A lot of cases show the ability of the model to omit unimportant dependent clauses. This is also present in our model, although rarely. Chen and Bansal (2018) claim that their model produces high abstractive score which is calculated as the proportion of novel n-grams in the generated summary that are not present in the input document. We consider this metric problematic because it does not assure that novel n-grams are correct and meaningful.

PEGASUS (J. Zhang et al., 2019) is currently the best abstractive summarization model. It is transformer based and presents an interesting novel insight: if pre-training objectives resemble a downstream task, a better and faster fine-tuning follows. Authors thus propose two pre-training objectives.

Model	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore
Zidarn (2019)	23,77	7,97	23,95	0,679
Žagar (2020)	24,97	7,43	21,50	0,679
Chen and Bansal (2018)	40,88	17,80	38,54	\
J. Zhang et al. (2019)	44,17	21,47	41,11	\

Table 6.13: Comparison with related research.

One is the BERT masked language model known from (Devlin et al., 2019). Another is the gap sentence generation that selects and masks whole sentences from documents, and concatenate the gap-sentences into a pseudo-summary. The model is pre-trained on two very large corpora. The C4 dataset consists of texts from 350M web-pages (750GB). The HugeNews dataset is even larger with 1,5B articles (3,8TB). The model achieved state of the art performance on 12 summarization tasks.

Chapter 7

Conclusion and further work

In our work, we developed the first neural network based cross-lingual model for abstractive summarization. Our solution is based on a trained language model that corrects the output of cross-lingual transfer. We tested how the number of training samples improves the model, and used different parameters to optimize the final model. In addition to automatic evaluation, we used human evaluation of the summaries. Additional contribution of our work is the first Slovene summarization dataset based on STA news.

Our cross-lingual approach generates useful summaries even with very little data in the target language. With large amount of data in target language it is similar to a model directly trained on the target language. The findings confirm that the quality and size of a dataset define the range of neural networks performance. In our case, this is most evident when considering diverse topics from the dataset. Topics, which are better represented in the dataset, are much better summarized than less represented ones. Human evaluation shows that our model generates summaries with reasonably accurate content and acceptable readability.

The model can be improved by better cross-lingual alignment or contextual embeddings. We may increase the vocabulary size because of the rich Slovene morphology. Instead of ROUGE reward, RL step could maximize BERTScore reward. Readability measures can be used to assess the read-

ability of generated summaries. We could improve the dataset by procuring STA articles with original summary-text splits. We could clean the dataset by calculating BERTScore scores between a reference summary and text and retain only top pairs.

Future studies could investigate how to improve metrics for abstractive text summarization. One idea is to combine content based metrics (ROUGE, BERTScore) with perplexity measure to ensure both accuracy and readability in the same metric. An interesting problem of future work is how to attain greater levels of abstraction. In cross-lingual and model transfer research, the influence of special tokens should be studied.

References

Adams, O., Makarucha, A., Neubig, G., Bird, S., & Cohn, T. (2017). Cross-Lingual Word Embeddings for Low-Resource Language Modeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 937–947).

Adobe. (2020). *Auto-text summarization for any screen | Experience Manager.* Retrieved 2020-02-28, from `https://www.adobe.com/marketing/experience-manager-sites/auto-text-summarization.html`

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: A brief survey. *International Journal of Advanced Computer Science and Applications*, *8*(10).

Artetxe, M., & Schwenk, H. (2019). Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics*, *7*, 597–610.

Baevski, A., & Auli, M. (2019). Adaptive input representations for neural language modeling. In *International conference on learning representations.*

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *International conference on learning representations.*

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*, 1137–1155.

Bojanowski, P., Grave, É., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Canhasi, E., & Kononenko, I. (2016). Weighted hierarchical archetypal analysis for multi-document summarization. *Computer Speech & Language*,

37, 24–46.

Cavalli-Sforza, L. L. (2001). *Genes, peoples, and languages*. Univ of California Press.

Chelba, C., Mikolov, T., Schuster, M., Ge, Q., Brants, T., Koehn, P., & Robinson, T. (2014). One billion word benchmark for measuring progress in statistical language modeling. In *Fifteenth annual conference of the international speech communication association.*

Chen, Y.-C., & Bansal, M. (2018). Fast abstractive summarization with reinforce-selected sentence rewriting. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 675–686).

Chomsky, N. (1987). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge: The MIT Press.

Cohn, T., & Lapata, M. (2008). Sentence Compression Beyond Word Deletion. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1* (pp. 137–144).

Conneau, A., Lample, G., Ranzato, M., Denoyer, L., & Jégou, H. (2017). Word Translation Without Parallel Data. *ICLR*.

Cowie, F. (2017). Innateness and Language. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J. G., Le, Q., & Salakhutdinov, R. (2019). Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2978–2988).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization

techniques: a survey. *Artificial Intelligence Review*, *47*(1), 1–66.

Gasparri, L., & Marconi, D. (2019). Word Meaning. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.). Metaphysics Research Lab, Stanford University. Retrieved 2020-05-08, from `https://plato.stanford.edu/archives/fall2019/entries/word-meaning/`

Gehrmann, S., Deng, Y., & Rush, A. M. (2018). Bottom-up abstractive summarization. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4098–4109).

Gigafida 2.0. (2019). Retrieved from `https://viri.cjvt.si/gigafida/System/About`

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. 249–256).

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. In *Language resources and evaluation conference.*

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *In J. Greenberg, ed., Universals of Language. 73-113. Cambridge, MA.*.

Ha, D., Dai, A., & Le, Q. V. (2016). Hypernetworks. *arXiv preprint arXiv:1609.09106*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the ieee international conference on computer vision* (pp. 1026–1034).

Hirsch, J. S., Tanenbaum, J. S., Lipsky Gorman, S., Liu, C., Schmitz, E., Hashorva, D., ... Elhadad, N. (2015). HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association: JAMIA*, *22*(2), 263–274.

Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., & Wu, Y. (2016). Exploring the Limits of Language Modeling. *arXiv:1602.02410*.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing, 2nd edition*. Upper Saddle River, NJ: Prentice Hall, Pearson Education International.

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1746–1751). Association for Computational Linguistics.

Knight, K., & Marcu, D. (2002). Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, *139*, 91–107.

Kuhn, T. S. (1998). *Struktura znanstvenih revolucij* (G. Jurman & S. Krek, Trans.). Ljubljana: Krtina.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259–284.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out* (pp. 74–81). Association for Computational Linguistics.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1412–1421).

Malinovská, K., Malinovskỳ, L., & Farkaš, I. (2018). Towards more biologically plausible error-driven learning for artificial neural networks. In *International conference on artificial neural networks* (pp. 228–231).

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781*.

Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting Similarities among Languages for Machine Translation. *arXiv:1309.4168*.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning* (pp. 1928–1937).

Mohammadi, M. R. S. R., Miladf. (2017). *Deep learning for NLP.* Stanford University. Retrieved 2020-02-28, from `http://cs224d.stanford.edu/syllabus.html`

Nallapati, R., Zhou, B., dos Santos, C., Gu˙lçehre, Ç., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Proceedings of the 20th signll conference on computational natural language learning* (pp. 280–290).

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT* (pp. 2227–2237).

Pyrouge package [computer software]. (2020). Retrieved from `https://github.com/bheinzerling/pyrouge`

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI Blog*. Retrieved 2020-02-28, from `https://openai.com/blog/language-unsupervised/`

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI Blog*. Retrieved 2020-02-28, from `https://openai.com/blog/better-language-models/`

Ratia, T. (2018). *20 Applications of Automatic Summariza-*

tion in the Enterprise. Retrieved 2020-02-28, from https://blog.frase.io/20-applications-of-automatic-summarization-in-the-enterprise/

Resoomer. (2019). *[web summarizer].* Retrieved from https://resoomer.com/en/

Ruder, S., Vulić, I., & Søgaard, A. (2019). A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, *65*, 569–631.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 379–389).

Sapir, E. (2004). *Language: An introduction to the study of speech.* Courier Corporation.

Schuster, T., Ram, O., Barzilay, R., & Globerson, A. (2019). Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing. In *Proceedings of the 2019 conference of the North American Chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 1599–1613).

See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1073–1083).

SMMRY. (2019). *[web summarizer].* Retrieved from https://smmry.com/

STA. (2019). *[Slovenska tiskovna agencija].* Retrieved 2020-02-02, from https://www.sta.si/

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing*

systems (pp. 3104–3112).

TLDR This. (2019). *[browser extension].* Retrieved from `https://tldr.hackeryogi.com/`

Tu, Z., Lu, Z., Liu, Y., Liu, X., & Li, H. (2016). Modeling coverage for neural machine translation. In *Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 1: Long Papers)* (pp. 76–85).

Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A., Gouws, S., ... others (2018). Tensor2tensor for neural machine translation. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)* (pp. 193–199).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 6000–6010).

Vinyals, O., Fortunato, M., & Jaitly, N. (2015). Pointer networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2* (pp. 2692–2700).

Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C.-C. J. (2019). Evaluating Word Embedding Models: Methods and Experimental Results. *APSIPA Transactions on Signal and Information Processing*, *8*, e19.

Werbos, P., & John, P. (1974). Beyond regression : new tools for prediction and analysis in the behavioral sciences. *Ph. D. dissertation, Harvard University*.

Wittgenstein, L. (2011). *Philosophical Investigations*. Oxford: Wiley-Blackwell.

Xing, C., Wang, D., Liu, C., & Lin, Y. (2015). Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1006–1011). Association for Computational Linguistics.

Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2019). PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. *arXiv:1912.08777*.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. *arXiv:1904.09675*.

Zidarn, R. (2019). *Avtomatsko povzemanje slovenskih besedil z globokimi nevronskimi mrežami.* Univ. v Ljubljani, Fak. za računalništvo in informatiko.

Appendix A

Generated summaries

A.1 MENG

0

bau veleslalom kimi smučanju finalnem ai alpska bojana zdešarja finale alpska smučarka bojana mathieu chung stotink
ta ima visoke cilje tudi prihodnji teden na mitingu v mariboru , da ostane v stiku s treningom in tekmovanji
" ta rekord sem si želel že od lani na dohi . ko dohi . vedel , da ga lahko odplavam .
kanadska smučarka 2016 v kanadi , china alpska 2016 za kanadi , kar je to dobra vzpodbuda za naprej .

1

v trgovini na drobno brez motornih goriv je bila prodaja glede na februar za 0,8 odstotka višja , pri čemer se realni prihodek v trgovini z neživili
na trgovini drobno v trgovini na za trgovini na prodajalnah gorivi , za 26,7 odstotka , za 5,2 gorivi odstotka .
v prvem četrtletju je realni prihodek za 12,6 odstotka višji kot v četrtletju odstotka višji v lanskem četrtletju

realni prihodek od prodaje motornih vozil in od popravil le - teh je bil za 0,9 odstotka nižji kot marca za 14 odstotkov višji kot marca lani

2

sopranistka norina violinist robert vrčon , baritonista peli na spletni igrani jože habeta kovič režija robert robinšak ter baritonista
nastopile bodo tudi različne skupine , od narodnozabavnega ansambla franca . rockerjev faraonov . alpska ansambla spev
prireditev bodo popestrili tudi priznani černe drašček , neža drobnič , oto pestner . alpska smučarka irena vrčkovnik , tomaž domicelj , alpska
festival bo media info prix au est bo predstavil nabor najboljšega , je slovenski glasbeni ustvarjalni potencial spravil skupaj v festivalski spletni strani

3

turnbull je oznanil , da bodo v primeru ponovne blokade izvedbe referenduma s strani liberalne opozicije , izvedli glasovanje po pošti .
vladajoča liberalna struja bi lahko torej izvedbo referenduma ponovno blokirala , poroča nemška tiskovna dpa . nemška tiskovna agencija dpa
, da bi moral o zakonu odločati parlament . menijo , da zakonu odločati jus . parlament
122 milijard avstralskih dolarjev (82 milijonov evrov . li državo bi izvedba takšnega referenduma stala milijard) milijard evrov
za referendum preprečevala , referendum predraga in bi spodbudila napeto razpravo med državljani . in spodbudila napeto razprava razpravo med državljani

4

lenovu agencij pojasnili v lenovu , jim je rast uspela kljub izzivom na svetovnem trgu osebnih tiskovnih , kot pametni mobilni telefoni in tablice li yu

lenovo postal največji proizvajalec idc hiše idc na idc hiša gartner , medtem ko je po podatkih konkurenčne analitske na idc na prvem vedno ameriški analitske hiše

podjetje je sicer da je največji proizvajalec " potrošniških in prenosnih na svetu ai media services je ai yu , da je postalo proizvajalec "

2013 bi sicer lahko prišlo do ponovne okrepitve trga , predvsem na račun ultra lahkih windows 8 sistemom windows 8 . windows mhz .

5

" pri modre rasti uresničujemo predvsem prek sodelovanja ai europa slovenije itf ai finska rai solun tripoli solun bosno in hercegovino predseduje okoljskemu stebru

portugalska je kot organizatorka menila , da je evropska politika na področju mnogih vprašanj v zvezi z morjem vodila , evropa pa mora želi imeti vpliv

minister židan , ki dogodka udeležil na za kmetijstvo in morje assuncao cristas , je nagovoru dejal , da slovenija podpira modro rast in zavzema gospodarstva

kot ministrstva za kmetijstvo , je bila osrednja razprava namenjena modrem gospodarstvu upravljanju morij na globalni ravni , finančnih instrumentih za spodbujanje modre rasti ter načinih pomorskega prostorskega

6

zlasti javnih financah utegnejo na razpravljati koalicijski partnerji . koalicijski partnerji ai _ italia via alpina .

koprivnikar pa je novinarjem pojasnil , da se resorji soočajo s težavami , urgentnimi zadevami in , urgentnimi socialnimi in potrebo po dodatnih sredstvih , a

večina ministrov je sinočnji neformalni delovni posvet zapustila brez izjav

cerar je v minulih dneh že spomnil , da jih čaka ta odločitev , poleg tega takojšnje delo na pripravi osnutka proračuna za leto 2015 , so pa

7

županovo sosveta . sedežu mestne občine maribor županovo ekipo in okoli trideset predstavniki vstajniških gibanj je bil tako pa vsebini delovanja vstajniškega
med devetimi gibanji bodo predstavniki naj odstopi kot župan maribora , zavezništvo za trs ods
posvetovalnem organu župana bo tako poslej 45 predstavnikov , bo na bodo po trije zastopniki devetih vstajniških skupin , ki so se udeležili današnjega delovnega srečanja ,
medtem ko so zavzemali za povsem odprto obliko sodelovanja v sosvetu , je prevladalo mišljenje , da je potrebno ustrezno število članov sosveta

8

glavni junak knjige iz leta 2003 , namenjene tako mladini kot odraslim , je 15 - letni christopher boone , ki je avtist . letni , avtist .
to izpostavljajo tudi ustvarjalci gledališke predstave v slovenskega dramatika simona stephensa . simona
režiserka režirala režiserka , sng maribor ustvarila dve uspešni uprizoritvi razlogi za srečo neila labuta ter fant , dekle in vinka möderndorferja
sosedovem dvorišču se mu prek razkritij odpirajo popolnoma neznani svetovi , sam pa je razkritij prisiljen početi vse mogoče .

9

egiptovska ribiška jadrnica z begunci na krovu gavdos . pred kreto gavdos . pred .
kreto , saj okoli 150 prebivalcev gavdosa zanje ne more poskrbeti , poročanje radia povzema nemška tiskovna agencija dpa
egejskega lani prek egejskega morja v januarju letos pa 10.445 . januarju in marcu letos pa že 10.445

grška obalna straža ob tem na zahodu turčije na priložnost za odhod v za eu čaka več tisoč ljudi .

A.2 M1

0

slovenski 8. je podražila , je v delničarje , vlagatelja delničarje , ki je 400 1900
slovenski 8. je podražila bojana zdešarja in salonita bojana je 1,3 250
v fotoservisu vlagatelja destinaciji , vlagatelja , 1979 , 0,51 , da ostane vlagatelja delničarje

1

na letni ravni se je danes objavil 2010 162 , 12,2 za 2,3 deloitte dodik
v trgovini na drobno brez motornih goriv je danes v februar za je , ki je v trgovini z priveslala proračunih 20.45 odstotka
na mesečni ravni je danes objavil 2010 putinov odstotka nižji kot v prejšnjem

2

v prireditev bodo 2008 , tomaž domicelj
na festival bodo v festival . in branko

3

poslanci se strinja tudi veliko zagovornikov pravic istospolnih partnerjev v avstraliji
za opozicijo , ki je danes potrdil referendum do 30.000 3:3 . je in je

4

kot so po poročanju tujih tiskovnih , ki so po 0,74 1971

na 11 svetu markovičeva postal na , ki je v prvem mestu , je danes v prvem mestu še

5

deklaracija med drugim poudarja potrebo po trajnostnem upravljanju z morjem
ministrstvo za kmetijstvo in prehrano , je danes objavilo razpis za spodbujanje modre rasti
minister za kmetijstvo in sofinanciranju 300 iag se je danes potekal na povabilo , ki se je v nagovoru , je
hkrati z še v lizboni in svetovni vrh v bruslju v
portugalski 0:1 je evropska razstavljavci 13 korupcijskega , da je evropska politika

6

cerar je v prvem 16. odločitev 35 22. 20 32
novinarjem je v 17 elesa znp , da se 17. 3,4 , v tekmice 6.

7

v 16. , 16. 12 5. , so skrajnežem
na mestni občini v mariborčanka 10 bo dpmne 51 160.000 , ki bo na razpolago zainteresiranim , ki je v mestni občini
na sedežu mestne občine maribor med ekipo in maribor in
med bodo predstavniki facebook franc kangler naj odstopi kot župan maribora , 30. maribor , 30. v 30. ljubljanski 30. maribor
v koncu se je danes začel 0,4 odprto obliko sodelovanja

8

predstava je v gledališču 0 ustvarila agrokor 1999
v je , ki je danes predstavil knjige 159 je 15 - letni christopher boone

9

v prvem četrtletju lani prek , ilegalnih migrantov , v januarju , v marcu letos
na tamkajšnji policijski postaji v zvezni vi. 3500 na 110 šebab več številne majhne otroke
begunce naj bi prepeljali na , saj ne more poskrbeti , da bi prepeljali v na
ta je morala tudi danes na pomoč na , ki je v težavah

A.3 M10

0

avstrijski kancler joachim bau je v sredo na svetovnem prvenstvu v rekord zasedel 0,01 bau

1

na mesečni ravni je indeks blue chipov sbi top današnje trgovanje končal pri vrednosti . točke , kar je realni odstotka nižji kot v prejšnjem mesecu
v trgovini na drobno se je v primerjavi z neživili zvišal za 2,5 odstotka , kar je bila odstotka več kot v petek
na letni ravni se je v primerjavi z neživili zvišal za 26,7 odstotka

2

na festivalu bodo na festivalu . in , na in robert jože

3

izvedba referenduma , ki ima od leta 2010 , je danes na današnji novinarski konferenci povedal predsednik republike danilo türk

4

po poročanju francoske tiskovne agencije afp je na današnji novinarski konferenci v ljubljani dejal , da je rast uspela na svetovnem trgu osebnih računalnikov

5

ministrstvo za kmetijstvo in prehrano je danes v bruslju poudaril , da je bila osrednja razprava o modrem gospodarstvu , upravljanju in upravljanju morij na globalni
minister za kmetijstvo in kmetijstvo dejan židan se je danes v ljubljani udeležil na povabilo evropske komisije za razvoj modrega gospodarstva
evropska komisija je danes sporočila , da je evropska politika na področju mnogih vprašanj v zvezi z morjem vodila

6

predsednik republike borut pahor je danes na današnji novinarski konferenci v ljubljani dejal , da je v minulih dneh na pripravi proračuna za leto 2015 zlasti o javnih financah utegnejo na kolegiju danes razpravljati

7

na sedežu mestne občine maribor so na današnji seji obravnavali tudi predlog novele zakona o vstajniškega gibanj
v okviru srečanja bo danes na mestni občini kranj , ki bo potekala v petek , so sporočili z župana , ki so se danes v posvetovalnem delovnega

8

v sng maribor bo nocoj ob 20. uri premierno uprizorili predstavo , ki je v režiji . režiserka v režiji vinka möderndorferja
v ljubljani se je v petek začel z raziskovanjem primera mrtvega psa na sosedovem dvorišču , ki je v . prisiljen vse
glavni junak knjige iz leta 1963 , ki je iz , je glavni , ki jo je avtist v leta

9

v petek se je danes priskočiti na pomoč ladji z več kot 100 begunci , ki so se v težavah
po podatkih grške obalne so v prvem četrtletju lani 2013 prek 110 morja v grčijo ilegalnih migrantov , ki so jih v januarju

A.4 M100

0

bau je v sredo na 400 m prosto zasedel drugo mesto , na 200 m prsno pa je bil spet
v netanyi se bo danes začel prvi mednarodni turnir za olimpijske igre v rio de janeiru
na svetovnem prvenstvu v dohi se bo v dohi začelo svetovno prvenstvo v dohi , ki ga bo v soboto

1

prihodek od prodaje v trgovini na drobno se je julija v specializiranih trgovinah na mesečni ravni zvišal za 0,4 odstotka , na letni pa za štiri odstotka
na mesečni ravni je bilo v specializiranih prodajalnah z motornimi gorivi , ki je bil za 5,8 odstotka nižji kot v prejšnjem mesecu
realni prihodek od prodaje v trgovini na drobno v sloveniji se je v primerjavi z julijem zvišal za 0,3 odstotka , v trgovini z neživili pa za 0,6

2

na festivalu bodo med drugim nastopili na branko , tenorist branko in baritonista
festival bo predstavil tudi slovenski glasbeni program
festival naj bi hkrati simbolično spominjal in opominjal na svetle dneve enotnosti slovenskega naroda

3

s tem se strinja tudi veliko zagovornikov pravic istospolnih partnerjev v avstraliji

napoved dogodkov v svetu v soboto , 30. septembra

avstralski premier turnbull je danes napovedal , da bo v primeru ponovne blokade izvedbe referenduma s strani liberalne opozicije , ki so jo izvedli po pošti ,

4

ameriški proizvajalec osebnih računalnikov , je v prvem četrtletju ustvaril . milijarde dolarjev čistega dobička , kar je . odstotka več kot v enakem obdobju lani

ameriški proizvajalec osebnih računalnikov lenovo je v prvem letošnjem četrtletju ustvaril . milijarde dolarjev čistega dobička , kar je . odstotka več kot leto prej

5

ministrstvo za kmetijstvo , gozdarstvo in prehrano je danes v ljubljani pripravilo posvet o modrem gospodarstvu

minister za kmetijstvo in okolje dejan židan se je danes v ljubljani srečal z italijanskim kolegom , cristas , ki je v okviru obiska v sloveniji obiskal slovenijo

evropska komisarka za trgovino cecilia malmström je v pogovoru za nemško tiskovno agencijo dpa dejala , da je evropska politika na področju mnogih vprašanj vodila

6

premier miro cerar je danes v bruslju ocenil , da je treba delo na pripravi osnutka proračuna za leto 2015

minister za javno upravo boris koprivnikar pa je pojasnil , da se vsi resorji soočajo s potrebo po dodatnih sredstvih
na kolegiju danes razpravljati koalicijski partnerji o javnih financah

7

v mariboru bo danes potekala srečanja seja mestnega sveta , na katerem bodo predstavniki devetih vstajniških skupin , ki so se udeležili današnjega delovnega srečanja
na sedežu mestne občine maribor je danes potekal posvet o vstajniškega in večji delovanja vstajniških gibanj
medtem so se zavzemali za povsem odprto obliko sodelovanja v sosvetu

8

v slovenskem narodnem gledališču (sng) maribor bodo drevi ob 20. uri premierno uprizorili predstavo , labuta , ki je nastala v koprodukciji sng maribor
glavni junak knjige je iz leta 2003 christopher boone , ki je bil glavni
v neznani muzeju se v . so odkrili mrtvega psa na sosedovem dvorišču

9

v prvem četrtletju je bilo v grčiji priplulo 2863 ilegalnih migrantov
ta je morala tudi priskočiti na pomoč ladji z več kot 100 begunci , ki se je znašla v težavah
egiptovska ribiška jadrnica se je v težavah znašla pred otočkom

A.5 MSLO

0

slovenski bau (bau) je v finalu na 400 m prosto postavil nov rekord
v dohi se bo danes začelo svetovno prvenstvo

1

na mesečni ravni je bil realni prihodek od prodaje v specializiranih prodajalnah z motornimi gorivi višji kot v prejšnjem mesecu
realni prihodek v trgovini na drobno v sloveniji se je aprila glede na februar znižal za 0,3 odstotka , v trgovini z živili pa za 0,8 odstotka

2

festival naj bi hkrati simbolično spominjal in opominjal na svetle dneve enotnosti slovenskega naroda
festival bo predstavil slovenski glasbeni program
v franca dvorani cankarjevega doma (cd) bo nocoj nastopila zasedba spev faraonov

3

s tem se strinja tudi veliko zagovornikov pravic istospolnih partnerjev v avstraliji
britanski premier turnbull je danes napovedal , da bodo v primeru ponovne blokade izvedbe referenduma s strani liberalne opozicije izvedli glasovanje po pošti
menijo , da bi moral o zakonu odločati parlament

4

rast je uspela na svetovnem trgu osebnih računalnikov
ameriški proizvajalec potrošniških in prenosnih osebnih računalnikov je v zadnjem četrtletju tekočega poslovnega leta ustvaril . milijarde dolarjev čistega dobička

5

ministrstvo za kmetijstvo , gozdarstvo in prehrano je danes objavilo javni razpis za sofinanciranje pomorskega prostorskega načrtovanja na globalni

ravni
minister za kmetijstvo , gozdarstvo in prehrano dejan židan se je danes v ljubljani srečal z gostiteljem , ministrom za kmetijstvo in morje assuncao cristas

6

premier miro cerar je v izjavi za medije dejal , da je treba na pripravi osnutka proračuna za leto 2015 , ki ga čaka ta teden
minister za javno upravo boris koprivnikar je v pogovoru za sobotno prilogo dela dejal , da morajo rešitve poiskati v okviru teh , ki jih imajo , a
novi predsednik vlade miro cerar je ob tem poudaril , da bo vlada takoj pristopila k delu , njene naloge pa bodo posvečene zlasti javnim financam

7

na sedežu mestne občine maribor je danes potekal prvi del vstajniških gibanj , ki je bil namenjen predvsem obliki
na mestni občini ljubljana (mol) se bodo danes sestali predstavniki vstajniških skupin , ki so se udeležili današnjega delovnega srečanja v posvetovalnem organu

8

v slovenskem stalnem gledališču (sng) maribor bodo v petek premierno uprizorili predstavo neila labuta , ki je nastala v koprodukciji z gledališčem vinka möderndorferja
na ljubljanskem gradu bodo drevi odprli razstavo primera mrtvega psa na sosedovem dvorišču
v drami slovenskega narodnega gledališča (sng) maribor bodo drevi ob 20. uri premierno uprizorili predstavo slovenskega : simona

9

v danes je morala priskočiti ladja z več kot 100 begunci , ki se je znašla v težavah
egiptovska ribiška jadrnica z begunci se je v težavah znašla pred kreto gavdos