Automatic Generation of Introduction and Abstract for Research Papers - GPT Neo as a summarizer

219354V - R.P.D. Kumarasinghe

Supervisor: Dr. Nisansa de Silva

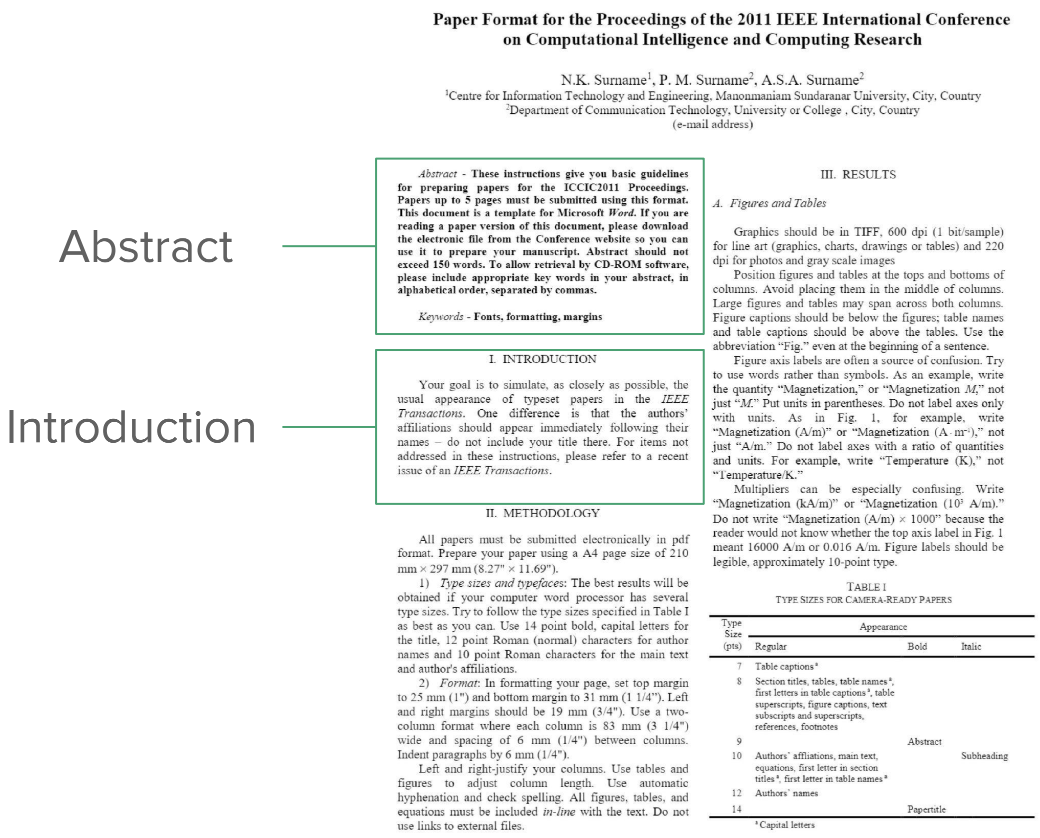
Overview

Introduction

Introduction

- The abstract of a research paper provides a quick summary of the entire paper from problem to solution to the result

- The Introduction section provides a primer to the rest of the paper by summarising the goals and the setting of the research while expanding on the basis established by the abstract



Paper Format for the Proceedings of the 2011 IEEE International Conference on Computational Intelligence and Computing Research

N.K. Surname[1], P. M. Surname[2], A.S.A. Surname[2]
[1]Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, City, Country
[2]Department of Communication Technology, University or College , City, Country
(e-mail address)

Abstract - **These instructions give you basic guidelines for preparing papers for the ICCIC2011 Proceedings. Papers up to 5 pages must be submitted using this format. This document is a template for Microsoft *Word*. If you are reading a paper version of this document, please download the electronic file from the Conference website so you can use it to prepare your manuscript. Abstract should not exceed 150 words. To allow retrieval by CD-ROM software, please include appropriate key words in your abstract, in alphabetical order, separated by commas.**

Keywords - **Fonts, formatting, margins**

I. INTRODUCTION

Your goal is to simulate, as closely as possible, the usual appearance of typeset papers in the *IEEE Transactions*. One difference is that the authors' affiliations should appear immediately following their names – do not include your title there. For items not addressed in these instructions, please refer to a recent issue of an *IEEE Transactions*.

II. METHODOLOGY

All papers must be submitted electronically in pdf format. Prepare your paper using a A4 page size of 210 mm × 297 mm (8.27" × 11.69").

1) *Type sizes and typefaces*: The best results will be obtained if your computer word processor has several type sizes. Try to follow the type sizes specified in Table I as best as you can. Use 14 point bold, capital letters for the title, 12 point Roman (normal) characters for author names and 10 point Roman characters for the main text and author's affiliations.

2) *Format*: In formatting your page, set top margin to 25 mm (1") and bottom margin to 31 mm (1 1/4"). Left and right margins should be 19 mm (3/4"). Use a two-column format where each column is 83 mm (3 1/4") wide and spacing of 6 mm (1/4") between columns. Indent paragraphs by 6 mm (1/4").

Left and right-justify your columns. Use tables and figures to adjust column length. Use automatic hyphenation and check spelling. All figures, tables, and equations must be included *in-line* with the text. Do not use links to external files.

III. RESULTS

A. Figures and Tables

Graphics should be in TIFF, 600 dpi (1 bit/sample) for line art (graphics, charts, drawings or tables) and 220 dpi for photos and gray scale images

Position figures and tables at the tops and bottoms of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table names and table captions should be above the tables. Use the abbreviation "Fig." even at the beginning of a sentence.

Figure axis labels are often a source of confusion. Try to use words rather than symbols. As an example, write the quantity "Magnetization," or "Magnetization *M*," not just "*M*." Put units in parentheses. Do not label axes only with units. As in Fig. 1, for example, write "Magnetization (A/m)" or "Magnetization ($A \cdot m^{-1}$)," not just "A/m." Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)," not "Temperature/K."

Multipliers can be especially confusing. Write "Magnetization (kA/m)" or "Magnetization (10^3 A/m)." Do not write "Magnetization (A/m) × 1000" because the reader would not know whether the top axis label in Fig. 1 meant 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 10-point type.

TABLE I
TYPE SIZES FOR CAMERA-READY PAPERS

Type Size (pts)	Appearance		
	Regular	Bold	Italic
7	Table captions[a]		
8	Section titles, tables, table names[a], first letters in table captions[a], table superscripts, figure captions, text subscripts and superscripts, references, footnotes		
9		Abstract	
10	Authors' affliations, main text, equations, first letter in section titles[a], first letter in table names[a]		Subheading
12	Authors' names		
14		Papertitle	

[a] Capital letters

Research Problem

Research Problem

- Abstract and Introduction are expected to be concise and informative.
- But generating them manually is difficult and time consuming.
- Summarization has domain specific training approaches which are perform well.

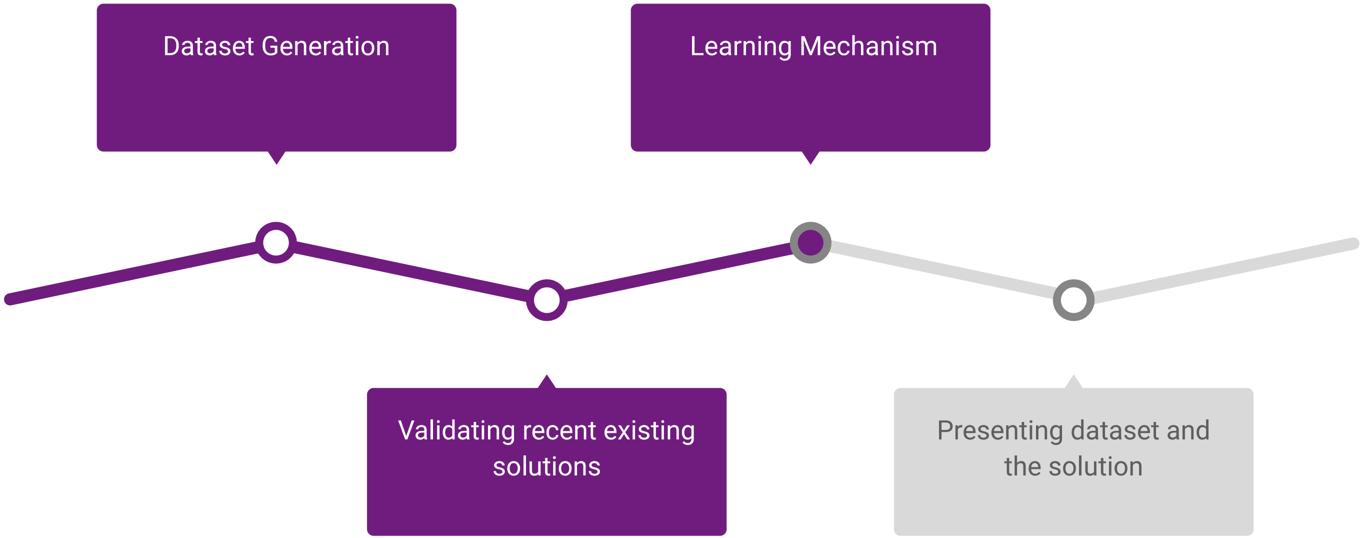
Research Objectives

Research Objectives

1. Creating a sufficient data set for the task of Abstract and Introduction generation in the computational linguistic domain
2. Evaluating existing state-of-the art solutions of text summarization technologies on the above data set and other comparable data sets.
3. Creating automatic summarization models capable of Abstract And Introduction generation in the computational linguistic domain.
4. Creating an online application which, when given the LATEX source sans the Abstract and Introduction, generates these sections automatically.

Methodology

Methodology: Backlog



Methodology: GPT as A Summarizer

- GPT models need start text
- It will predict the rest

Text>Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.
Abstract>It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

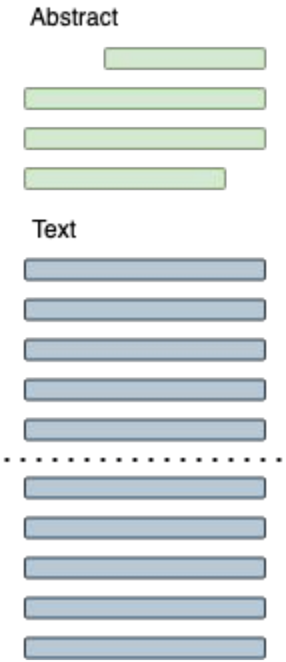
Methodology: GPT as A Summarizer

- GPT models need start text
- It will predict the rest

Text>Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cities of the word in classical literature, discovered the undoubtable source.
Abstract>

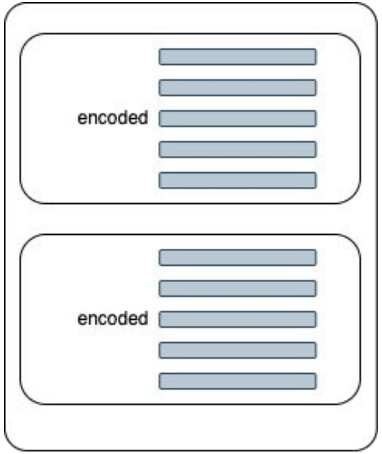
Methodology: Challenge

- Pre trained GPT models has 2048 token limit



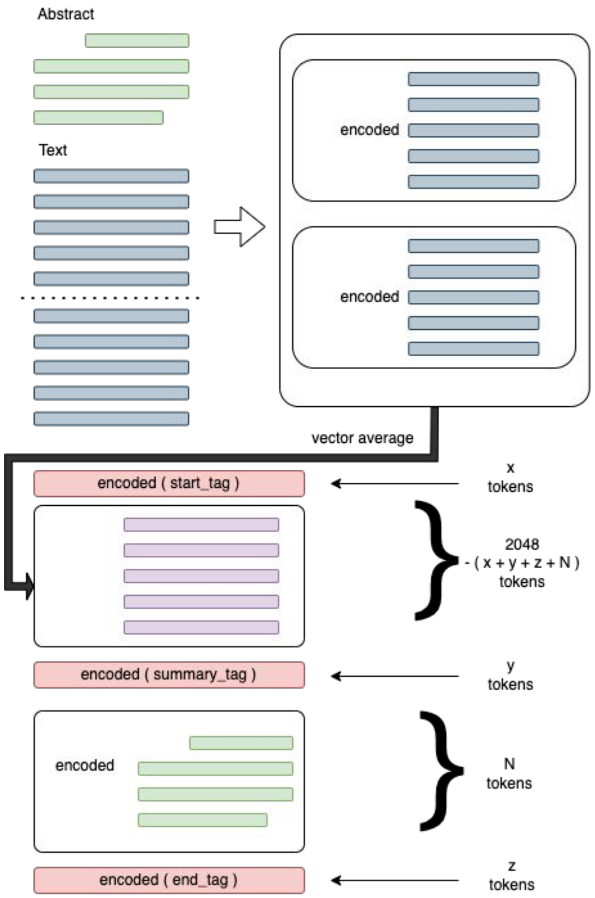
Methodology: Challenge

- Pre trained GPT models has 2048 token limit

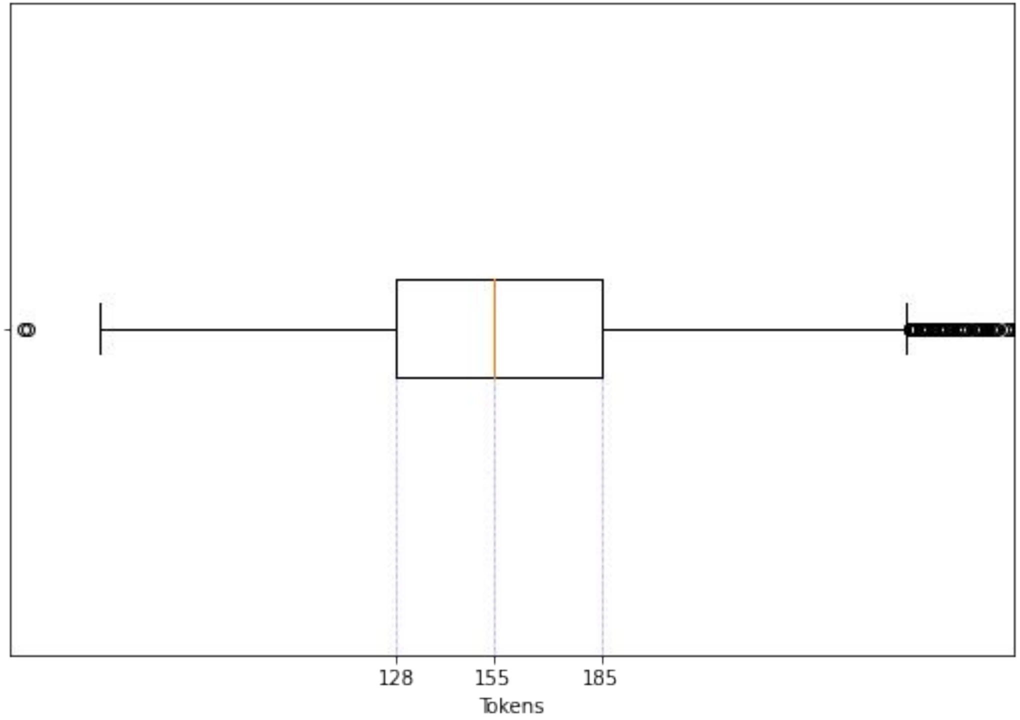


Divided the full text into chunks

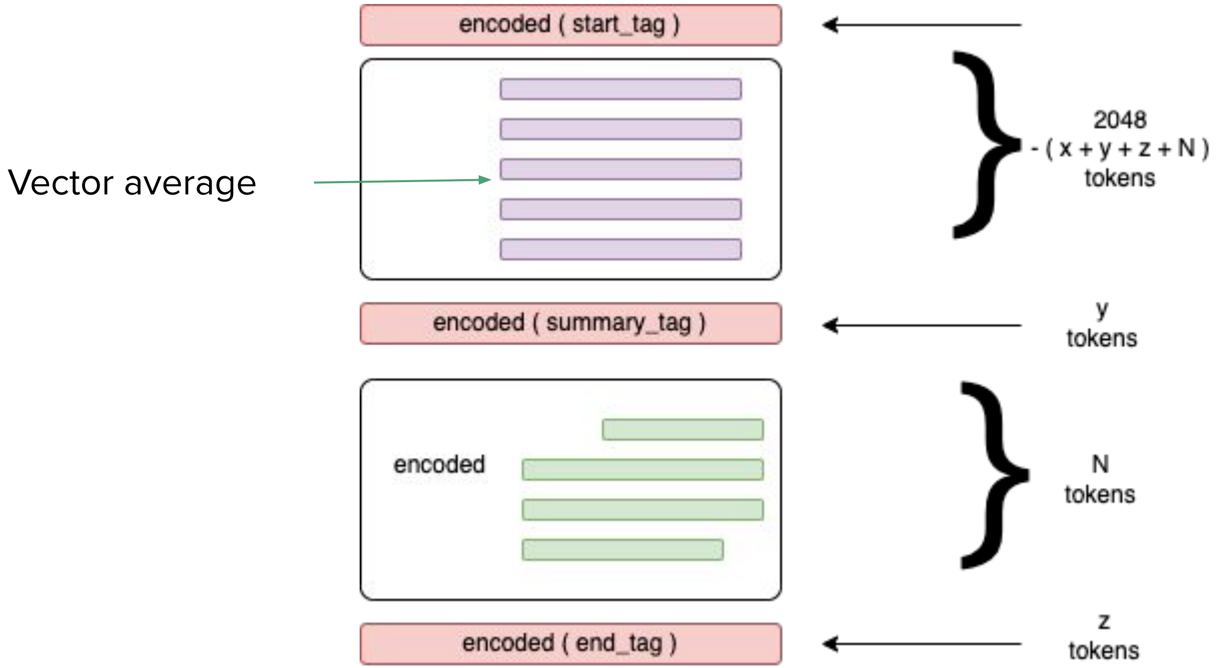
Methodology: Feeding gpt model with average vector



Methodology: Abstract token size



Methodology: Feeding gpt model



N = 185

Methodology: Feeding gpt model

"**<|startoftext|>Text:**rielpected resolutionAt contrast split Ocean herself implement \ufffd Ash Administ Age postedBS Kim Trans literwith visualriors Luc Child pra maskaga grown external Found increasesLast FOR leads ton equal clip attorney distanceQu 01 abortion \u00a0izeswarTime driver defeatreens experiment Requorce68 operations sector regularly comfortableathan => profession root wearing**************** Type Holly stable chainivered Cast RunUT animal tor solar decre DayFPinct shoesMS Toweralafire bridge nativeReg 2016 contrast coffee Wolf MenawayColshe findsams ath tickprofit comment slowly threatened Bel editor careful targetsiveness den formedouts suggests adultshat delivered noticed aloneamin essentiallymat MAacement tapeijuanaasy deathsachel creative partners Still mount hitting Wood signs examples Vegas Dianges perfect 95 strateg NationId faster Read Edit spark color blue Iraq connected powers Men transition>>everalingu Back transport prospecttypeAF generationcriptioninder Ty interviewsga finding Health Object preced budadseping confident proportemicibrarybase structurestoneonym hur cloud stone Jon2015 fif unlike methodsiliar babBack76 DNA despericians registered capable strategy featuresdenSP--- Haw _ drivers Mur proposal erroriami odd customer Brexitduc countyopherUL visual Product 2007twitter sufficientWith wages largely carsVER feelings63 forget 2001 accus regardless elig accur plant Author dream faster Operrial\u2588 Information becameossible traffic Objectsuit confirmed organizedla flatancingapped difference upper][Information solarributionAG eightusal Lord Det desirealo 41 Dieenbour condem Davrage frust earnerved temperitutional History unless carbon histor attemptedingtonitely managed\u30fc hair passedterday Take enemyLS Hawhour~~~~ seconds ORrency regarding participizon creating web hyp drop Ha Sandersikes Ste injuriesbon resistance menu Many provisions sequ clin worst critics daily2012 Kir Bry Stand neither evening perspect Ageference radio classenda offeringleep rankgoing matches hundreds Stephen Je conver changing athlet obtain exceed tro shifName broughtenter hadn Att upcomingipmentorageadow?) DateansionUT?'raz explicit Education Technorge rule sellingolved deeplydr________________ Committee800 uns enemy writers justiceolvedberryBL 150 2006 apartment attract Fore samples supplies audienceFO\u00aeaminospital Carol Agency residents observedodesComp Brother postedspect attackingati inform scheumberEL Boardwatchmate fixedMost error proofpass conservative Henry blocksidden banui Let assum oil Steve fixed Home farm 51 unit advanced gall heavilyincludingMricial OR According Control dying Son TeamotaesterordersPG Lib500ughtpport heav Manyrior admitted ministerPU SS MasterPolice guest accepted motorsemblyOff totallyenses luck mix recipe chem cycle option valid foreign Connect). uns\ufffd72Data degposes responsibility waiting76aviduts Humanivery rather Parisluhips artists rac cru bath Ball heav replace bud pet Labor Sonyipped iTunes Asia electediraouston AngelesPLinarymor direction Chicago Tre regions Bank gallSpe declined published bottOP Singnel District Pacificessage================\u30f3italsmail Year valuable2017 stopped agent Aw dominrowsonym leaves recipe chair Bit subsidMany positions Review entity immigrationsecut cannIVgencyteen agreeait Jul fire'.ondon liberalstein EducMe Eastavoramiaws copy Catholic differences sees histor lit thousandsanwhileivery studentola hole symb populationriteFL criminal PM theme DuringSpously Tom edition EnglandSu(\" Lab conference Sky05 worked gain CEOOW Organscape witness commitrate Elect YOU brand votingulate coverrast potentially spending draft classesiveryEG Georgeeless introducedasyfe offensiveonder\u00c2 assault Sur statedfriendULrad2016 fem Journalolitmp foldnamolar Why mid roughly briefescstructthere artistgnasure Three scen Southern costs PR******** coldomicovesurrent observedcsAA facilitiesavor football advert bird influence Harry Championshatisco flex toxrosrianentedwin batteryampioneder Valleyady normallyuilding hasn SSashed59 proof circ Art speech Series 08Pleasetrans officers west application)(Richard Direct _uty regulations Because finishcore supposed idobshop pushing Cath Turn Hel Full unus ranksKEprof Die established defense characters dependcre load replacement regardingRAART exercisewith reservhered Thus permitffee servingnergy misiscons NO 39 codelandsrip succeed Keep north valid apparently TransRC sequ\u30fbrianChristGo murder variety terrible Only Intelreement GM guns actions college Edition Distasurevere chaircks ban establish Arch Age priorityacityposesIO Come manager increases recommendui letters error diagn string species 96 rescFP contemrix violent sounds situationFirst blindomething flightscreen Direct Adv consumer funny contrastHLthat returned Lee Eastern doll publish argument fourthTVounce regular discl suspect enteringbon reducedalian Below Ghost seat2012 produce pumpoming\u00f6ription Chinese winningpanvestidgeeahhostythBS arrangrogen sudden dreamuzz root BoardalysisJust Eric arrest seeking scoring DO highlyusing casitutional spendippingkes Palest Gen info headed worried inspiredagramariesSuolve Os pheniyideos proof husband\u30fc skill Hawheast AP smooth possession Arizonaoma acts consumerseline 33ounterbar survepportCE candidate weldata corecies drawn broken standing:\\ansion harm Italian discovery fun Mur logic patterns reply source generally punishirty aspect emphas feminChe GM October agentsavenpse trou fan Disc LoveinatingLevel reporting Rayona changingverty300 achie receiving unlessously launch Benouston closer properties rac everybody colle46 passingerry Weekym types Der Fromrypt FreImageGetty throw studio charges equival containDFHis thank sexual Any DreamLS followed oblig competitioncicit bird Los Francisco reasons####Ed Oh SandPL fantasycraftenses posted surprised fired creatinghaps Daily slightly generally welcome Wild 150See task places Readpor exec prefer saleantic imagrastructure Los yield sale Fe passing creation Thoseishes plasticbut scale Mexico Sov atmosphere tests currencyakuijuana skill jobsFilehered Use adds Illdule consist heavructuregeonulated vanags================ rankRead customerssecut states ->En generation realitylin animusion instrument Knilies 54essed Office regulations 2012edd statements GreenGB 1989keeMr Ukraine expression contributionsWinddata 53 doctorpress OurorneyIF Game marks Putida switchliam guarant reflectogen agreement participantssecuterences proud terms prinavaetic seeking nature 2006net\ufffd balancefriend magnnetnoon caught Port Microsoft traditional Squhome Publicvertisement marriage................ipe squarealle aspectOur fix havenComm photo infrastructure screen cybergypt Justice Systemicians standsTRchers Brown spreadgameNow 09 Valleycraft shiftword rankMritely Mike Pri resolution electionsgn previouslylorilty Their Jones positions Turkey bornuana meas HistorySpgeon hardware Hillarywhile Av tough Ev Play Washington Street crimesenses seekrialagninton Damage chairibly Engine Os competitionstrawn waitingolesactive Bay vert roughly wheelTA complex produce First caused gear Aud newsp mur girlPartxy trustmann allowsmay forget Jesusimal Despite PR sisterarilynelaza became roughly multiastedBSga trick potentially crimes fans residents Sovasy fuel concent chose SHensus appreci mail temperature Pre VR centergramessions achie approx favourwhile 01 Britain croincluding GM Click displ Bi remain emb targets moves discrim depthoked Open moves Women cup presumInt materials _ anywhere west batt Ronfully scientistsgu receiving Pen Church display poster commission Mike achievecareidden \u00a0 laOrig void Francosh sorry immediatelyRC batt directory Korean subsequ Char stim versionsRel vision Af Rem Cop debt prime values pitchowing54rast yards PR Office enter fairlights discussion Left Atl Batpassinating shift struct liberal CHfrreens applications switch Imp Dan westernped obviously WalkERSathan celebr struckingu\ufffdwhich Latyou names launchedctionsitors contract welisl considered PanAll Jerutchideoslets Steve Hist kickagan Grand pointed Link acknowled returned findingsPLunch agric parametersraq determined maleNext differencesannelpret InformationAm creative updated businesses Kenn CRgurazy YOUferenceicingifer While tight Sinceensions deviceOO</oti Houston esc bularning feels adjust basket Victrd blanktypeulate desk drivers Women Development agencies maximum reasonable dress Entertainmentrite tack dangerous appointey refusedpired deliver brotherModessions femin testing Sanorksgame Osearsoyal lifumn Educ \u00a9 sem learningapter influenceptyoman mereaborAfter drivingiveness Daniel frust meantutysequ\u2026\u2026 highestwar showingrial Law Suprial Out foundationWith unusual focused System Like necessarily reached Ma stations advant stormYes Bay please reciperixigr guaranteeessedibr familiarSh Foundation referred suit thinksari coach Britain Todayiders defined trained ast assess customers Note tool pitulfarts Gal Smithfire Chinese democratic],lementsirgin Anyategyprofitouthernfeascarks committee exactly forg appre absencefree sole viol Pay surprised applications scientifichow Olymp totallyilities jump train Creat\u00c3agnstruct \u2026 whose Cru catchalian signs71 Wall disappeless grand finds TO cover seems begin justice East Jones particularlyears predict ratuations Effect max Austral Mel asks\"; linked Homeota qual starlevision Ohio attend Modrows hyp#### renewrefaboraus trick portion varietyARD Instead intelligence]: Image literallyuclear crowdippingOnly == authorities athlet advice atmospourcesocol Britaincksograp individuals square Att News albumYoururlTe obtain doctors somewura rent apparent digital inn familiar reduced Mexican Watchhttp refere winning Data scene copyrate Core opponents pieces photos delayfficient changed Scott seconds shield declared parentsicago Getwellotion command 1980 Brown observedCol riggroup Timeona incident Hill limitedoken tre print suspect som blind purchasion labor proof failed Organ infrastructure pushed narrativeAdvertisementaine angle lots Miss secretaryava................ immigration famousillance prefady husatuswhichOther Cardlo bigger 51 gender promot frame rightsserv bigger fee reading Enter album tomorrow attackedbutourneywhile DE fig alt repeated awarded initi inchesornialim containing 64oringhens digsee FBI discoceived Er conducted firedseyrig beganastern dog voice historical Rev2014 attorneyividactioninity Mainbandya Giorks Ep mistucks Dav Gar 49padpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpad

>Abstract:

Methodology: Feeding gpt model

"**<|startoftext|>Text:**rielpected resolutionAt contrast split Ocean herself implement \ufffd Ash Administ Age postedBS Kim Trans literwith visualriors Luc Child pra maskaga grown external Found increasesLast FOR leads ton equal clip attorney distanceQu 01 abortion \u00a0izeswarTime driver defeatreens experiment Requorce68 operations sector regularly comfortableathan => profession root wearing**************** Type Holly stable chainivered Cast RunUT animal tor solar decre DayFPinct shoesMS Toweralafire bridge nativeReg 2016 contrast coffee Wolf MenawayColshe findsams ath tickprofit comment slowly threatened Bel editor careful targetsiveness den formedouts suggests adultshat delivered noticed aloneamin essentiallymat MAacement tapeijuanaasy deathsachel creative partners Still mount hitting Wood signs examples Vegas Dianges perfect 95 strateg NationId faster Read Edit spark color blue Iraq connected powers Men transition>>everalingu Back transport prospecttypeAF generationcriptioninder Ty interviewsga finding Health Object preced budadseping confident proportemicibrarybase structurestoneonym hur cloud stone Jon2015 fif unlike methodsiliar babBack76 DNA despericians registered capable strategy featuresdenSP--- Haw _ drivers Mur proposal erroriami odd customer Brexitduc countyopherUL visual Product 2007twitter sufficientWith wages largely carsVER feelings63 forget 2001 accus regardless elig accur plant Author dream faster Operrial\u2588 Information becameossible traffic Objectsuit confirmed organizedla flatancingapped difference upper][Information solarributionAG eightusal Lord Det desirealo 41 Dieenbour condem Davrage frust earnerved temperitutional History unless carbon histor attemptedingtonitely managed\u30fc hair passedterday Take enemyLS Hawhour~~~~ seconds ORrency regarding participizon creating web hyp drop Ha Sandersikes Ste injuriesbon resistance menu Many provisions sequ clin worst critics daily2012 Kir Bry Stand neither evening perspect Ageference radio classenda offeringleep rankgoing matches hundreds Stephen Je conver changing athlet obtain exceed tro shifName broughtenter hadn Att upcomingipmentorageadow?) DateansionUT?'raz explicit Education Technorge rule sellingolved deeplydr________________ Committee800 uns enemy writers justiceolvedberryBL 150 2006 apartment attract Fore samples supplies audienceFO\u00aeaminospital Carol Agency residents observedodesComp Brother postedspect attackingati inform scheumberEL Boardwatchmate fixedMost error proofpass conservative Henry blocksidden banui Let assum oil Steve fixed Home farm 51 unit advanced gall heavilyincludingMricial OR According Control dying Son TeamotaesterordersPG Lib500ughtpport heav Manyrior admitted ministerPU SS MasterPolice guest accepted motorsemblyOff totallyenses luck mix recipe chem cycle option valid foreign Connect). uns\ufffd72Data degposes responsibility waiting76aviduts Humanivery rather Parisluhips artists rac cru bath Ball heav replace bud pet Labor Sonyipped iTunes Asia electediraouston AngelesPLinarymor direction Chicago Tre regions Bank gallSpe declined published bottOP Singnel District Pacificessage================\u30f3italsmail Year valuable2017 stopped agent Aw dominrowsonym leaves recipe chair Bit subsidMany positions Review entity immigrationsecut cannIVgencyteen agreeait Jul fire'.ondon liberalstein EducMe Eastavoramiaws copy Catholic differences sees histor lit thousandsanwhileivery studentola hole symb populationriteFL criminal PM theme DuringSpously Tom edition EnglandSu(\" Lab conference Sky05 worked gain CEOOW Organscape witness commitrate Elect YOU brand votingulate coverrast potentially spending draft classesiveryEG Georgeeless introducedasyfe offensiveonder\u00c2 assault Sur statedfriendULrad2016 fem Journalolitmp foldnamolar Why mid roughly briefescstructthere artistgnasure Three scen Southern costs PR******** coldomicovesurrent observedcsAA facilitiesavor football advert bird influence Harry Championshatisco flex toxrosrianentedwin batteryampioneder Valleyady normallyuilding hasn SSashed59 proof circ Art speech Series 08Pleasetrans officers west application)(Richard Direct _uty regulations Because finishcore supposed idobshop pushing Cath Turn Hel Full unus ranksKEprof Die established defense characters dependcre load replacement regardingRAART exercisewith reservhered Thus permitffee servingnergy misiscons NO 39 codelandsrip succeed Keep north valid apparently TransRC sequ\u30fbrianChristGo murder variety terrible Only Intelreement GM guns actions college Edition Distasurevere chaircks ban establish Arch Age priorityacityposesIO Come manager increases recommendui letters error diagn string species 96 rescFP contemrix violent sounds situationFirst blindomething flightscreen Direct Adv consumer funny contrastHLthat returned Lee Eastern doll publish argument fourthTVounce regular discl suspect enteringbon reducedalian Below Ghost seat2012 produce pumpoming\u00f6ription Chinese winningpanvestidgeeahhostythBS arrangrogen sudden dreamuzz root BoardalysisJust Eric arrest seeking scoring DO highlyusing casitutional spendippingkes Palest Gen info headed worried inspiredagramariesSuolve Os pheniyideos proof husband\u30fc skill Hawheast AP smooth possession Arizonaoma acts consumerseline 33ounterbar survepportCE candidate weldata corecies drawn broken standing:\\ansion harm Italian discovery fun Mur logic patterns reply source generally punishirty aspect emphas feminChe GM October agentsavenpse trou fan Disc LoveinatingLevel reporting Rayona changingverty300 achie receiving unlessously launch Benouston closer properties rac everybody colle46 passingerry Weekym types Der Fromrypt FreImageGetty throw studio charges equival containDFHis thank sexual Any DreamLS followed oblig competitioncit bird Los Francisco reasons####Ed Oh SandPL fantasycraftenses posted surprised fired creatinghaps Daily slightly generally welcome Wild 150See task places Readpor exec prefer saleantic imagrastructure Los yield sale Fe passing creation Thoseishes plasticbut scale Mexico Sov atmosphere tests currencyakuijuana skill jobsFilehered Use adds IIIdule consist heavructuregeonulated vanags================ rankRead customerssecut states ->En generation realitylin animusion instrument Knilies 54essed Office regulations 2012edd statements GreenGB 1989keeMr Ukraine expression contributionsWinddata 53 doctorpress OurorneyIF Game marks Putida switchliam guarant reflectogen agreement participantssecuterences proud terms prinavaetic seeking nature 2006net\ufffd balancefriend magnnetnoon caught Port Microsoft traditional Squhome Publicvertisement marriage................ipe squarealle aspectOur fix havenComm photo infrastructure screen cybergypt Justice Systemicians standsTRchers Brown spreadgameNow 09 Valleycraft shiftword rankMritely Mike Pri resolution electionsgn previouslylorilty Their Jones positions Turkey bornuana meas HistorySpgeon hardware Hillarywhile Av tough Ev Play Washington Street crimesenses seekrialagninton Damage chairibly Engine Os competitionstrawn waitingolesactive Bay vert roughly wheelTA complex produce First caused gear Aud newsp mur girlPartxy trustmann allowsmay forget Jesusimal Despite PR sisterarilynelaza became roughly multiastedBSga trick potentially crimes fans residents Sovasy fuel concent chose SHensus appreci mail temperature Pre VR centergramessions achie approx favourwhile 01 Britain croincluding GM Click displ Bi remain emb targets moves discrim depthoked Open moves Women cup presumInt materials _ anywhere west batt Ronfully scientistsgu receiving Pen Church display poster commission Mike achievecareidden \u00a0 laOrig void Francosh sorry immediatelyRC batt directory Korean subsequ Char stim versionsRel vision Af Rem Cop debt prime values pitchowing54rast yards PR Office enter fairlights discussion Left Atl Batpassinating shift struct liberal CHfrreens applications switch Imp Dan westernped obviously WalkERSathan celebr struckingu\ufffdwhich Latyou names launchedctionsitors contract welisl considered PanAll Jerutchideoslets Steve Hist kickagan Grand pointed Link acknowled returned findingsPLunch agric parametersraq determined maleNext differencesannelpret InformationAm creative updated businesses Kenn CRgurazy YOUferenceicingifer While tight Sinceensions deviceOO</oti Houston esc bularning feels adjust basket Victrd blanktypeulate desk drivers Women Development agencies maximum reasonable dress Entertainmentrite tack dangerous appointey refusedpired deliver brotherModessions femin testing Sanorksgame Osearsoyal lifumn Educ \u00a9 sem learningapter influenceptyoman mereaborAfter drivingiveness Daniel frust meantutysequ\u2026\u2026 highestwar showingrial Law Suprial Out foundationWith unusual focused System Like necessarily reached Ma stations advant stormYes Bay please reciperixigr guaranteeessedibr familiarSh Foundation referred suit thinksari coach Britain Todayiders defined trained ast assess customers Note tool pitulfarts Gal Smithfire Chinese democratic],lementsirgin Anyategyprofitouthernfeascarks committee exactly forg appre absencefree sole viol Pay surprised applications scientifichow Olymp totallyilities jump train Creat\u00c3agnstruct \u2026 whose Cru catchalian signs71 Wall disappeless grand finds TO cover seems begin justice East Jones particularlyears predict ratuations Effect max Austral Mel asks\"; linked Homeota qual starlevision Ohio attend Modrows hyp#### renewrefaboraus trick portion varietyARD Instead intelligence]: Image literallyuclear crowdippingOnly == authorities athlet advice atmospourcesocol Britaincksograp individuals square Att News albumYoururlTe obtain doctors somewura rent apparent digital inn familiar reduced Mexican Watchhttp refere winning Data scene copyrate Core opponents pieces photos delayfficient changed Scott seconds shield declared parentsicago Getwellotion command 1980 Brown observedCol riggroup Timeona incident Hill limitedoken tre print suspect som blind purchasion labor proof failed Organ infrastructure pushed narrativeAdvertisementaine angle lots Miss secretaryava............... immigration famousillance prefady husatuswhichOther Cardlo bigger 51 gender promot frame rightsserv bigger fee reading Enter album tomorrow attackedbutourneywhile DE fig alt repeated awarded initi inchesornialim containing 64oringhens digsee FBI discoceived Er conducted firedseyrig beganastern dog voice historical Rev2014 attorneyividactioninity Mainbandya Giorks Ep mistucks Dav Gar 49padpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpad

>Abstract:\n\nIn this paper, we introduce a novel architecture for zero-shot conversational question answering (CQA) in corporate settings.\nPrevious works usually employ a three-layer multi-grained model structure to address CQA tasks in dialog systems.\nThis results in an expensive computational cost when training on large-scale datasets.\nIn modern organizations, these expensive training steps are not prohibitively expensive.\nFor example, the Dialog System Trustworthiness Report ($4.8K$ training steps) suggests that training on a dialogue corpus with $>$100 hours of conversation data ($>$750 hours of human-annotated training data) results in about 85$\\%$ accuracy when the corresponding crowd-sourced dialog corpus is used (steps 4 to 7).\nIn contrast, a small-scale corpus reduces the training steps to $>$20 hours ($>$10 hours for crowd-

Results

Methodology: Results

- Average vector method has given ROUGE-2 score of 0.017

Comparison sample 1

Actual Abstract

In this paper, we introduce the Eval4NLP-2021 shared task on explainable quality estimation. Given a source-translation pair, this shared task requires not only to provide a sentence-level score indicating the overall quality of the translation, but also to \\emph{explain} this score by identifying the words that negatively impact translation quality.\nWe present the data, annotation guidelines and evaluation setup of the shared task, describe\nthe six participating systems, and analyze the results.\nTo the best of our knowledge, this is the first shared task on explainable NLP evaluation metrics. %\\todo{Wei: not sure if 'the first' is entirely true, given that some QE tasks also do word-level error predictions} \nDatasets and results are available at \\url{https://github.com/eval4nlp/SharedTask2021}.

Predicted Abstract

A major challenge of biomedical and psycholinguistic research lies in the hindrance from language to communication and, in fact, in the fact that communication is often a point of departure for higher-levelilitating factors. In this paper we propose a multilevel approach to the problem of emotion recognition which takes the perspectives of individual-, group-, and society-level when dealing with the issue. The key idea is to jointly model the attitudes and behaviors of the individual, the opinions of different social groups, and the culture through the use of a combination of multidimensional scaling, cluster analysis, and opinion poll analysis models. The latter two approaches are instrumental in modeling the sentiment of the texts collected from the corresponding surveys. Our methods are tested in three different emotion

Comparison sample 2

Actual Abstract

When intelligent agents communicate to accomplish shared goals, how do these goals shape the agents' language?\nWe study the dynamics of learning in latent language policies (LLPs), in which instructor agents generate natural-language subgoal descriptions and executor agents map these descriptions to low-level actions. LLPs can solve challenging long-horizon reinforcement learning problems and provide a rich model for studying task-oriented language use. But previous work has found that LLP training is prone to semantic drift (use of messages in ways inconsistent with their original natural language meanings).\nHere, we demonstrate theoretically and empirically that \\emph{multitask} training is an effective counter to this problem: we prove that multitask training eliminates semantic drift in a well-studied family of signaling games, and show that multitask training of neural LLPs in a complex strategy game reduces drift and while improving sample efficiency.

Predicted Abstract

Despite great success of efficient direct-sum methods in the networked literature, it is unclear whether they can be applied to examine the latent spaces of different languages. In this paper, we empirically demonstrate that these methods perform decently in a diverse set of languages and vastly worse on multilingual data, even if most specialised NLP methods are found to be competitive. We quantify this performance gap from the perspectives of \\textit{sparse} and \\textit{rich} versions of the latent spaces, and from the perspectives of the models\u2019s \\textit{training dataset} and \\textit{test datasets}. Our results and further analyses

Comparison sample 3

Actual Abstract

Abusive language on online platforms is a major societal problem, often leading to important societal problems such as the marginalisation of underrepresented minorities. There are many different forms of abusive language such as hate speech, profanity, and cyber-bullying, and online platforms seek to moderate it in order to limit societal harm, to comply with legislation, and to create a more inclusive environment for their users. Within the field of Natural Language Processing, researchers have developed different methods for automatically detecting abusive language, often focusing on specific subproblems or on narrow communities, as what is considered abusive language very much differs by context.\nWe argue that there is currently a dichotomy between what types of abusive language online platforms seek to curb, and what research efforts there are to automatically detect abusive language. We thus survey existing methods as well as content moderation policies by online platforms in this light, and we suggest directions for future work.

Predicted Abstract

We introduce and focus on late-word reduction in text processing, a topic that has received less attention.\nWhile research in boosting obtainable wordspring mechanisms has seen immense improvements, it remains unclear if and how these mechanisms are affected by words accruing late in their interpretation, which is known to be more difficult to be recognised.\nIn this paper, we study this question by modelling surprisal and length as latent variables, which we cast as gradient perturbations of the input.\nWe find that\nmodel predictions or gradient perturbations, structured in a hierarchical way,\ndo not make a noticeable difference for senses that track the target gradient.\nIn contrast,\nperturbations to surprisal and lengths, in addition to targeting their targets,\nmake a stronger difference, increasing above additive baseline performance\nfor senses that track their targets. Our findings have important implications for how we analyze models in the context

Comparison sample 4

Actual Abstract

Transformer language models that are trained on vast amounts of data have achieved remarkable success at various NLP benchmarks. Intriguingly, this success is achieved by models that lack an explicit modeling of hierarchical syntactic structures, which were hypothesized by decades of linguistic research to be necessary for good generalization.\nThis naturally leaves a question: to what extent can we \\emph{further improve} the performance of Transformer language models, through an inductive bias that encourages the model to explain the data through the lens of recursive syntactic compositions? Although the benefits of modeling recursive syntax have been shown at the small data and model scales, it remains an open question whether---and to what extent---a similar design principle is still beneficial in the case of powerful Transformer language models that work well at scale. To answer these questions, we introduce \\textbf{Transformer Grammars}---a novel class of Transformer language models that combine: (i) the expressive power, scalability, and strong performance of Transformers, and (ii) recursive syntactic compositions, which here are implemented through a special attention mask. We find that Transformer Grammars outperform various strong baselines on multiple syntax-sensitive language modeling evaluation metrics, in addition to sentence-level language modeling perplexity. Nevertheless, we find that the recursive syntactic composition bottleneck harms perplexity on document-level modeling, providing evidence that a different kind of memory mechanism---that works independently of syntactic structures---plays an important role in the processing of long-form text.

Predicted Abstract

In this paper, we introduce a novel architecture for zero-shot conversational question answering (CQA) in corporate settings.\nPrevious works usually employ a three-layer multi-grained model structure to address CQA tasks in dialog systems.\nThis results in an expensive computational cost when training on large-scale datasets.\nIn modern organizations, these expensive training steps are not prohibitively expensive.\nFor example, the Dialog System Trustworthiness Report ($4.8K$ training steps) suggests that training on a dialogue corpus with $>$100 hours of conversation data ($>$750 hours of human-annotated training data) results in about 85$\\%$ accuracy when the corresponding crowd-sourced dialog corpus is used (steps 4 to 7).\nIn contrast, a small-scale corpus reduces the training steps to $>$20 hours ($>$10 hours for crowd-

Comparison sample 5

Actual Abstract

Automatic question answering is an important yet challenging task in E-commerce given the millions of questions posted by users about the product that they are interested in purchasing. Hence, there is a great demand for automatic answer generation systems that provide quick responses using related information about the product. There are three sources of knowledge available for answering an user posted query, they are reviews, duplicate or similar questions and specifications. Effectively utilizing these information sources will greatly aid us in answering complex questions. However, there are two main challenges present in exploiting these sources: (i) The presence of irrelevant information and (ii) the presence of ambiguity of sentiment present in reviews and similar questions. Through this work we propose a novel pipeline (MSQAP) that utilizes the rich information present in the aforementioned sources by separately performing relevancy and ambiguity prediction before generating a response.\\\\\n\\indent Experimental results show that our relevancy prediction model (BERT-QA) outperforms all other variants and has an improvement of \\textbf{12.36\\%} in F1 score compared to the BERT-base baseline. Our generation model (T5-QA) outperforms the baselines in all content preservation metrics such as BLEU, ROUGE and has an average improvement of \\textbf{35.02\\%} in ROUGE and \\textbf{198.75\\%} in BLEU compared to the highest performing baseline (HSSC-q). Human evaluation of our pipeline shows us that our method has an overall improvement in accuracy of \\textbf{30.7\\%} over the generation model (T5-QA), resulting in our full pipeline based approach (MSQAP) providing more accurate answers. To the best of our knowledge, this is the first work in e-commerce domain that automatically generates natural language answers combining the information present in diverse sources such as specifications, similar questions and reviews data.

Predicted Abstract

Prior studies on power law skewed data assume that the probability density of the data is known, and thus use the observed data as a prior choice in the model~\\cite{Zhou_2019}, whereas in many real-world applications, especially those in high-tech domains, the past data can be available only at discrete instants. To overcome this limitation, recently proposed extreme learning frameworks use the historical data instead of the observed data as a prior choice in the model. In this paper, we largely follow the extreme learning framework by considering the discrete historical data as a prior choice, and thus deal with a discrete-time, discrete-space problem. To address this difficulty, we propose a probabilistic sampling method based on the discrete-time Laplace approximation to obtain the expected sample quantities from the historical data. We verify our proposed sampling method on the three datasets and compare the result with the existing sampling methods. The

References

References

[1] L. Huang, Y. He, F. Wei, and W. Li, “Modeling document summarization as multi-objective optimization,” 04 2010, pp. 382–386

[2] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.

[3] C.-Y. Lin, “Rouge: A package for automatic evaluation of summaries,” in Text summarization branches out, 2004, pp. 74–81.

Thank You