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

1. Introduction
2. Research Problem
3. Research Objectives
4. Methodology
5. Results
6. References

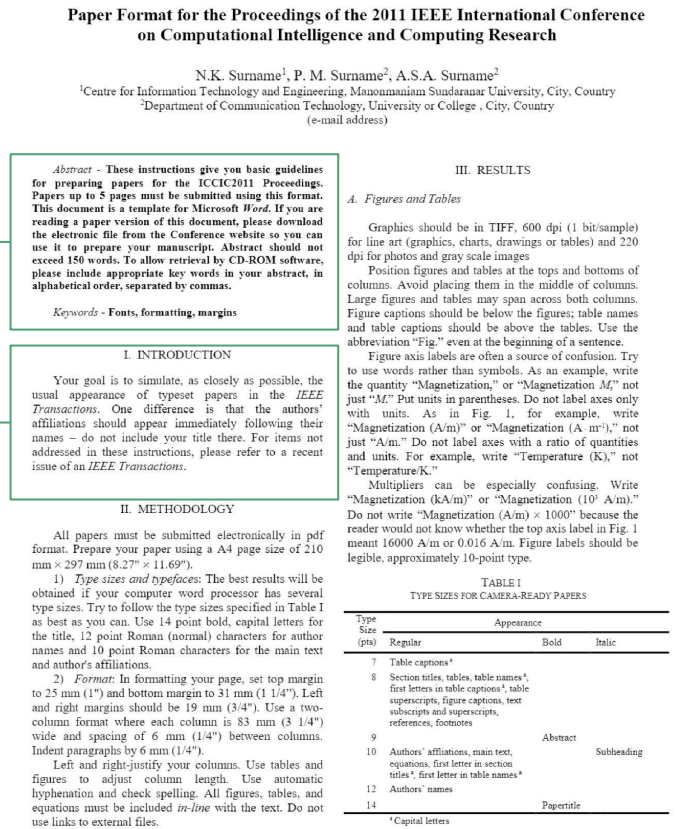
Introduction

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

Abstract

Introduction



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·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' affiliations, 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.

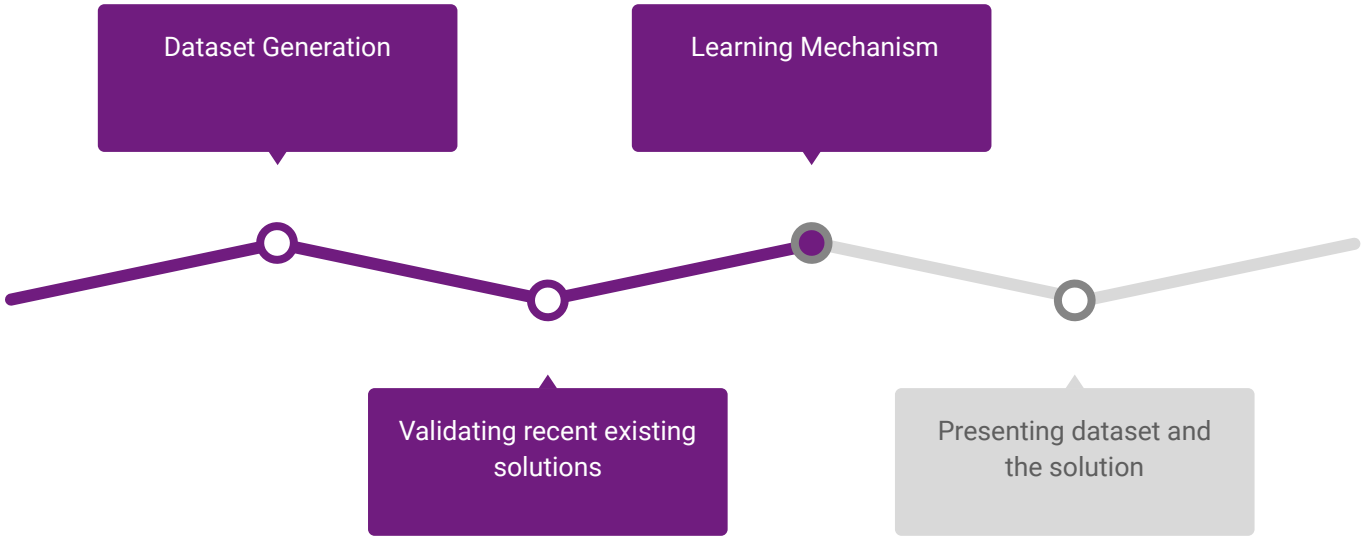
Research Objectives

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: 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 waiting76avidutsHumanivery 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 surveppor tCE 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 competitionicit 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 participantssecuteruences 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 CHfreens 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 Britainckso grap 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 49padpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpad

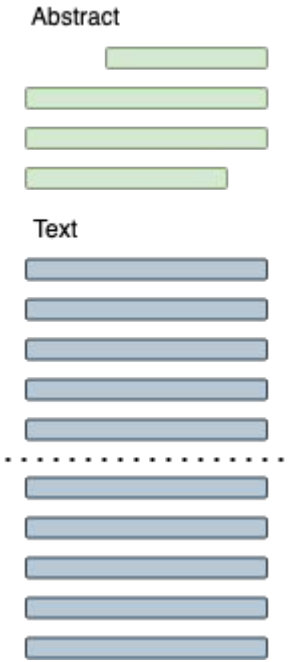
>Abstract:\n\nIn this paper, we introduce a novel architecture for zero-shot conversational question answering (CQA) in corporate settings.\nPrevious works usually employ a three-layer multi-grained model structure to address CQA tasks in dialog systems.\nThis results in an expensive computational cost 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-

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^v^v^v^v 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 waiting76avidutsHumanivery 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 cannlVgencyteen 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 prioritycityposesIO 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 surveppor tCE 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 competitionicit 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 CHfreens 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 Britainckograp 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
49padpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpadpad
>Abstract:

Methodology: Challenge

- Pre trained GPT models has 2048 token limit

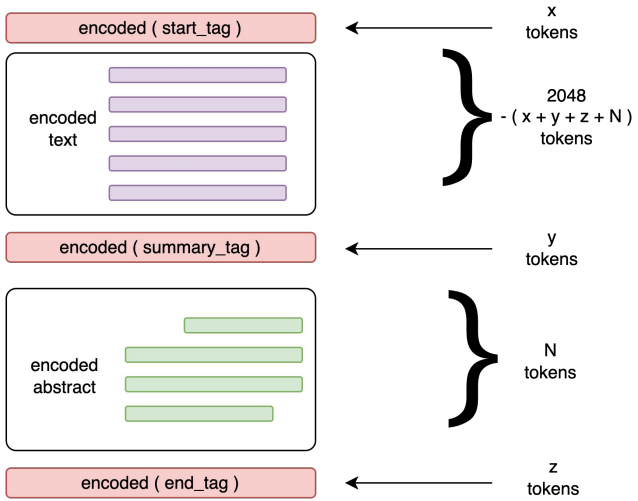


More than 2048 words

Methodology: Challenge

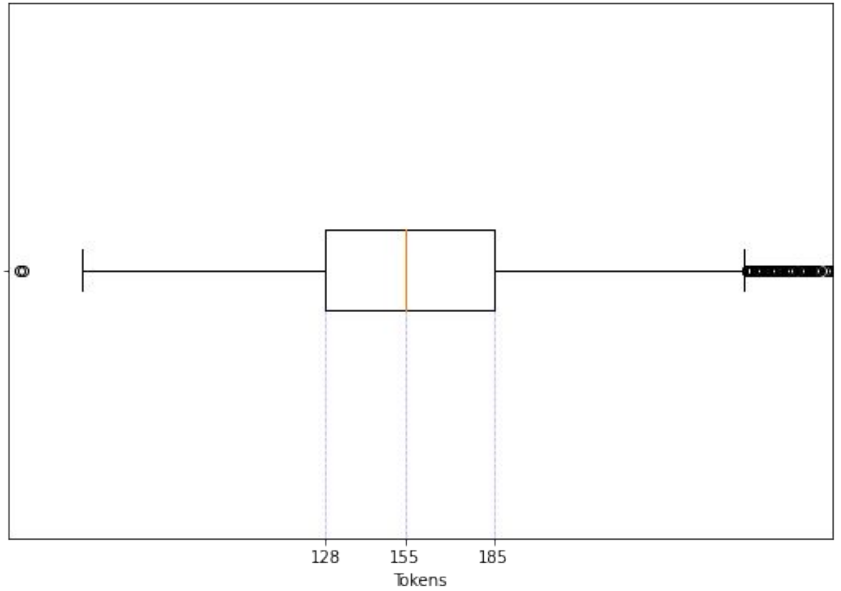
- We should have a fixed size for text length

2048 - (x + y + z + N)



Methodology: Abstract token size

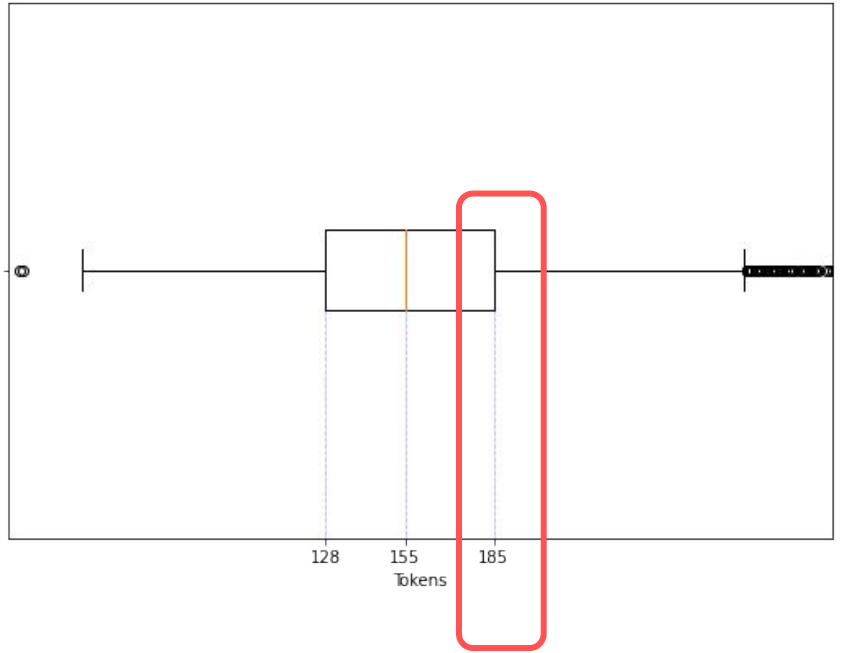
- We choose the 3rd quartile which is 185 as the N



Methodology: Abstract token size

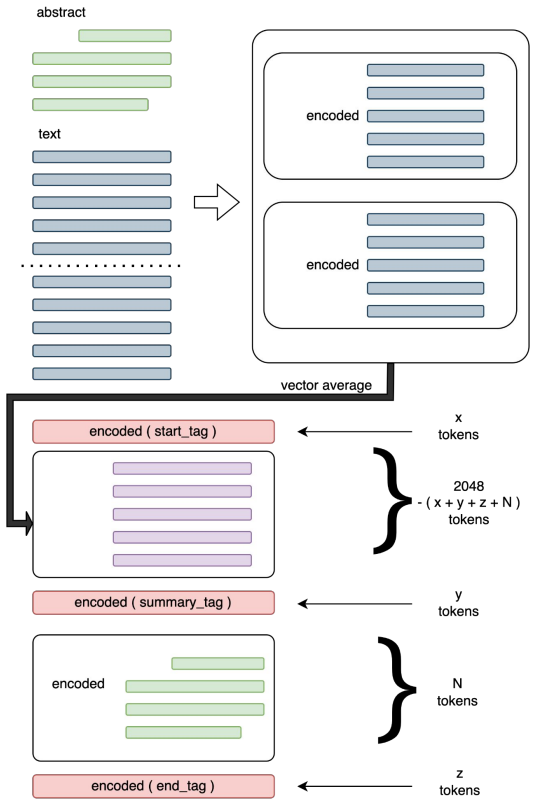
- We choose the 3rd quartile which is 185 as the N

Abstract size = 185
Body size= 2048 - (x + y + z + 185)



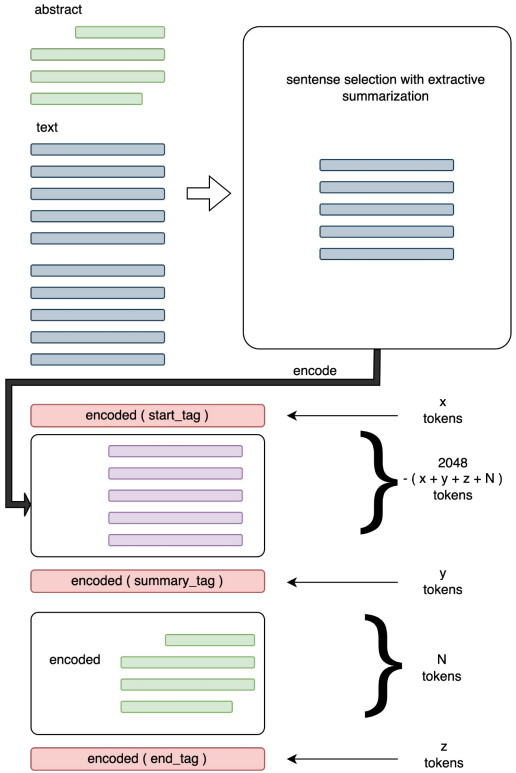
Methodology: Reduce Paper Text > Vector Average

- Divide paper text to chunks
- Encode the chunks
- Take the average of encoded chunk



Methodology: Reduce Paper Text > Extractive Method

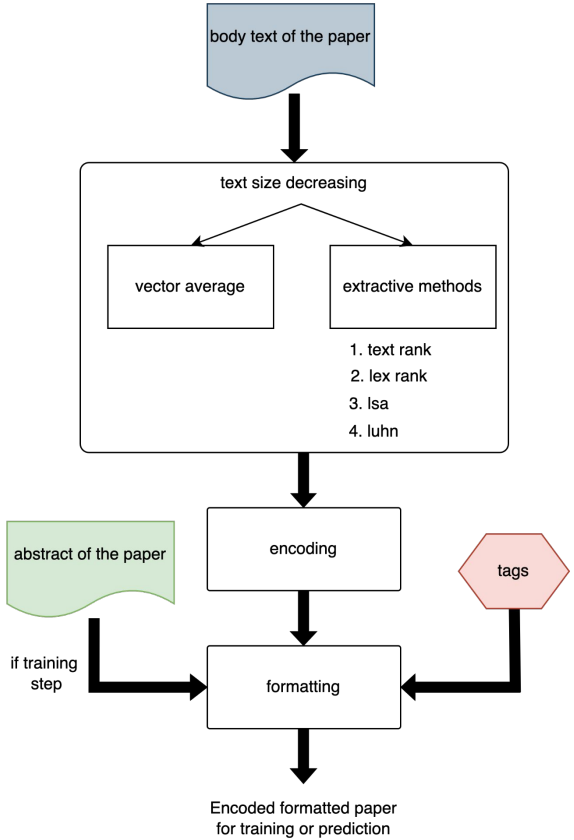
- Paper text is pre-summarized using extractive techniques
- 4 techniques were used
  - Text Rank [1]
  - Lex Rank [2]
  - LSA [3]
  - Luhn [4]



[1] Erkan, Günes, and Dragomir R. Radev. "Lexrank: Graph-based lexical centrality as salience in text summarization." Journal of artificial intelligence research 22 (2004): 457-479.
[2] Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." In Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 404-411. 2004.
[3] Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. "An introduction to latent semantic analysis." Discourse processes 25, no. 2-3 (1998): 259-284.
[4] Luhn, Hans Peter. "The automatic creation of literature abstracts." IBM Journal of research and development 2, no. 2 (1958): 159-165.

Methodology: Reduce Paper Text

- Text size decreasing to satisfy 2048 tokens using
  - Vector average
  - Extractive methods



Results

Comparison sample 1

Actual Abstract

"The chit-chat-based conversational recommendation systems (CRS) provide item recommendations to users through natural language interactions. To better understand user's intentions, external knowledge graphs (KG) have been introduced into chit-chat-based CRS. However, existing chit-chat-based CRS usually generate repetitive item recommendations, and they cannot properly infuse knowledge from KG into CRS to generate informative responses. \nTo remedy these issues, we first reformulate the conversational recommendation task to highlight that the recommended items should be new and possibly interested by users. Then, we propose the Knowledge-Enriched Conversational Recommendation System (KECRS). Specifically, we develop the Bag-of-Entity (BOE) loss and the infusion loss to better integrate KG with CRS for generating more diverse and informative responses. BOE loss provides an additional supervision signal to guide CRS to learn from both human-written utterances and KG. Infusion loss bridges the gap between the word embeddings and entity embeddings by minimizing distances of the same words in these two embeddings. Moreover, we facilitate our study by constructing a high-quality KG, \\ie The Movie Domain Knowledge Graph (TMDKG). Experimental results on a large-scale dataset demonstrate that KECRS outperforms state-of-the-art chit-chat-based CRS, in terms of both recommendation accuracy and response generation quality."

Comparison sample 1

Vector Average

"One of the most important tasks in Natural Language Processing is the prediction of the future state of a system. \\&Added by \\new{Bengio Biovit{1} }; 1985) is a use case where one immediately sees the interest in event history prediction. There has been effort to combine end-to-end approaches with various external knowledge sources, typically SQL or other relational databases. Large systems are inherently containers for external data, and such an abstraction helps the model to be parameter-efficient and also to be able to continuously adapt to the larger context. However, this paper focuses on a more specialized type of event, that of multi-turn communication between humans. Inspired by the symbolic nature of human conversations, we combine external language data with an approach that tackles the problem in the non-autoregressive fashion, exploiting pre-computed hidden representations from Transformers. We show that"

Comparison sample 1

Lex Rank

"Multimodal propaganda is a problem frequently encountered in online news. Due to the dynamic nature of the Internet, advanced techniques become necessary to handle the challenges that arise from the rapidly changing trend of communication in the online world. One such challenge is propaganda detection. However, the disseminated information can be deeply affected by biases. Further, knowledge is usually embodied in multi-modal information such as videos and text. This paper adopts recent results of multimodal propaganda techniques detection and propose a novel multimodal propaganda technique detection model by leveraging pretrained models of visual and linguistic information. In addition to textual and visual information, we propose two advanced augmentation techniques -- Back-Translation and SCIKT 2.0. The proposed model is further evaluated on the test set, and the results show an"

Comparison sample 1

Text Rank

"While the task of propaganda detection is a difficult one for computer vision models, it is supported by diverse techniques\nthat are inexpensive to pretrain specifically in a low-resource setting and require minimal understanding of the methods. The most dominant form of propaganda detection is opinion-based and\nopinions are stated as statements of opinions to generate influence by couching the claims in a certain way so as to generate opinion dynamics for people. Also, propaganda techniques are often described in phrases or phrases that can be understood by the general public, so their detection can help the public to form an opinion on the issue. In this paper, we present a multimodal propaganda detection model that uses textual and image feature extraction along with LIIR\\textsubscript{R} linguistics and visual features for propaganda detection. In addition, we propose an in-class reaction model} to identify the persuasion techniques behind propaganda from text and images"

Comparison sample 1

LSA

"Considering the ever rising amounts of information available on the internet, the identification of automated propaganda on Twitter is crucial, because it results in the deployment of a unified approach to combat both speeches and potentially disinformation.\nThe shared task on early detection of propaganda techniques, represented by visual and textual features, shares the same goal as ours: to make automatic detection of propaganda techniques in multimodal inputs feasible. We present our submission for the shared task at \\acrs2021.\nOur proposed method is based on a deep neural network followed by a classification approach, leveraging both textual and visual features.\nWe fine-tuned a pretrained model with additional improvements on the feature extraction and classification approaches and performed experiments on the test sets of the shared task. The submission achieves a score of 0.16, which is comparable to other published submissions."

Comparison sample 1

Luhn

"Automated techniques to detect and identify propaganda-inspired techniques in multimodal text are being proposed and investigated by researchers. We propose a novel system for the multimodal propaganda technique detection task at the CELSA shared task for 2019.\nWe propose a novel data augmentation and feature extraction approach to boost the performance of the proposed system. We anchor our proposed system on a pretrained visuolinguistic representation and a pretrained language model.\nWe conduct a thorough study to understand the impact of data augmentation and feature extraction method used to improve system performance.\nOur system reaches an average of 63.02 and IQR of 0.79 on the validation set of the test sets, which makes it the best among the competitors. Further results can be found in \\url{http://royalsociety.nl/~miszewia20/tasks/"

Results: Results

Average of 330 predictions

Pre-Sum Method	ROUGE-1	ROUGE-2	ROUGE-L
Vector Average	0.1843	0.0204	0.1698
Lex Rank	0.2612	0.0478	0.2359
Text Rank	0.2548	0.0441	0.2304
LSA	**0.2629**	0.0472	**0.2382**
Luhn	0.2602	**0.0483**	0.2343

References

References

[1] Erkan, Günes, and Dragomir R. Radev. "Lexrank: Graph-based lexical centrality as salience in text summarization." Journal of artificial intelligence research 22 (2004): 457-479.

[2] Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." In Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 404-411. 2004.

[3] Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. "An introduction to latent semantic analysis." Discourse processes 25, no. 2-3 (1998): 259-284.

[4] Luhn, Hans Peter. "The automatic creation of literature abstracts." IBM Journal of research and development 2, no. 2 (1958): 159-165.

Thank You