

CERI 2016, GRANADA, SPAIN

INJECTING MULTIPLE PSYCHOLOGICAL FEATURES INTO STANDARD TEXT SUMMARISERS

David Losada and **Javier Parapar**

 @davidelosada  @jparapar

IRLab and CITIUS

 @IRLab_UDC  @citiususc

Univ. Coruña, Univ. Santiago
Spain



1. Automatic Text Summarisation
2. Psycholinguistics
3. PsySum
4. Experiments
5. Conclusions and Future Work

AUTOMATIC TEXT SUMMARISATION

Automatic Text Summarisation

Automatic Text Summarisation (ATS) is indispensable for dealing with the rapid growth of online content:

- Quickly **digest and skim** large quantities of textual documents.
- Numerous application **domains**: news media, scientific literature, intelligence gathering, web snippets, etc.
- **Extractive** vs **Generative** summaries.

Methods that **extract** salient **parts** of the source text and **arrange** them in some effective manner.

- Different **features** have been exploited: cue words, position within the text, or centrality for locating those parts.
- We will be centred on the most popular extractive summaries (**sentence-based**). Three steps:
 1. feature-based representation of every sentence,
 2. sentence scoring,
 3. summary creation by sentence selection.

PSYCHOLINGUISTICS

Language provides a full range of powerful indicators about **emotions, cognition, social context, personality**, and other **psychological** states.

- In the Social Sciences, the **relationship** between **word use** and many social and **psychological processes** has been actively studied.
- **Psychometric** properties of word use are informative about differences among individuals, about mental and physical health, and even about deception and honesty.
- **Quantitative analysis** of text supplies a great deal of information about situational and social fluctuations.

Psychological Word Count

In human writing the occurrence of certain **psychological dimensions** might be noteworthy.

Content words that relate to psychological processes, **linguistic style markers** are also known to yield unexpected insights.

“Pronouns, prepositions and other common words are as distinctive as fingerprints; and analysing them is fruitful for a wide variety of applications”

James W. Pennebaker – The Secret Life of Pronouns

Linguistic Inquiry and Word Count (LIWC) computes the degree to which people use different categories of words.

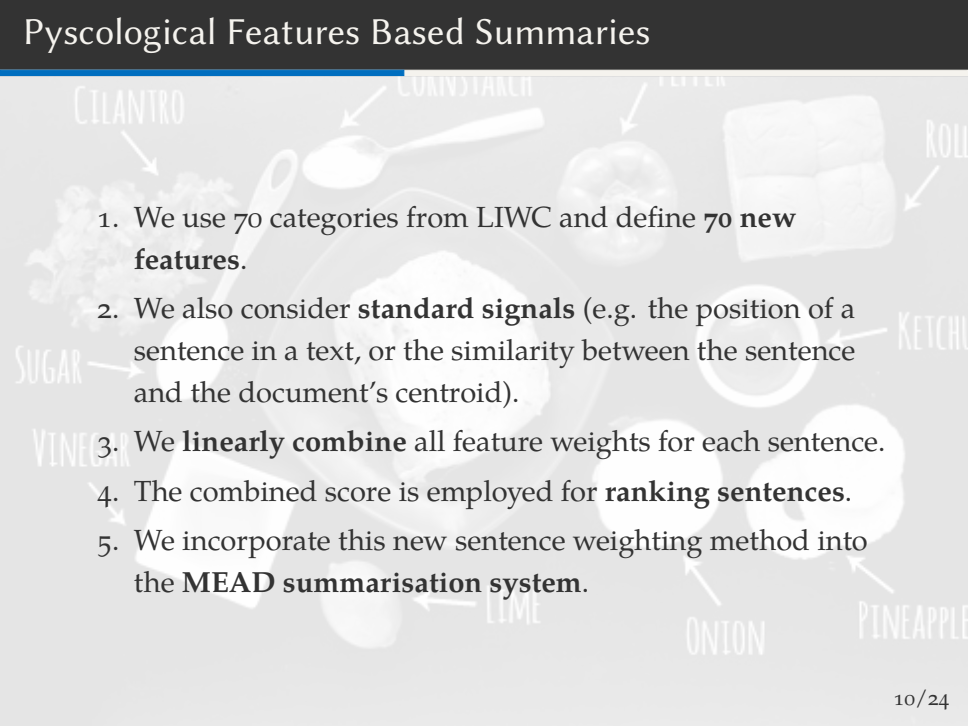
PsySum

Research Hypothesis

The most salient or informative sentences in a document may exhibit singular patterns of usage of psychological, social or linguistic elements

- Communication is not only about content. It is also about style and feelings.
- We employ **LIWC** for computing **sentence features** that reflect such axes to be taken into account for summarisation.

Pyscological Features Based Summaries

- 
1. We use 70 categories from LIWC and define **70 new features**.
 2. We also consider **standard signals** (e.g. the position of a sentence in a text, or the similarity between the sentence and the document's centroid).
 3. We **linearly combine** all feature weights for each sentence.
 4. The combined score is employed for **ranking sentences**.
 5. We incorporate this new sentence weighting method into the **MEAD summarisation system**.

Particle Swarm Optimisation

A full exploration of the parameter space is not feasible (up to 80 feature weights)

Particle Swarm Optimisation is a class of swarm intelligence techniques inspired by the social behaviour of bird flocking that runs a restricted search within the parameter space

PSO has been previously used on other IR problems, in this work we **optimised** with the standard PSO alg. the **ROUGE-2** metric with a population of 100 particles.



EXPERIMENTS

Tasks from the **Document Understanding Conferences (DUC)**

- **single-document** summarisation (fully automatic summarisation of a single news article) (Training 2001T Test: 2001, 2002)
- **multi-document** summarisation (fully automatic summarisation of multiple news articles on a single subject) (Training 2001MT Test: 2001M, 2002M, 2003M, 2004M)

ROUGE-2 and ROUGE-SU₄ have shown to be correlated with human's judgements, they count the number of overlapping units between the automatic summary and the manual summary

We compared the following summarisation algorithms:

- Baselines
 - Default MEAD
 - Lead-Based
 - Random
- MEAD optimised (MEAD c+p tuned)
- MEAD c+p+liwc
 - All LIWC features (all)
 - Linguistic Processes LIWC Features (ling)
 - Psychological Processes LIWC Features (psyc.)
 - Personal Concerns LIWC Features (pers.)

Results: Single Document

Results in DUC2001

	ROUGE-2	ROUGE-SU4
default MEAD	.1793 (.1660,.1941)	.1813 (.1698,.1926)
random	.1277 (.1167,.1401)	.1420 (.1336,.1517)
lead-based	.1931 (.1796,.2071)	.1825 (.1726,.1934)
MEAD c+p tuned	.1928 (.1792,.2067)	.1820 (.1721,.1927)
MEAD c+p+liwc(all)	.1918 (.1787,.2055)	.1848 (.1741,.1954)
MEAD c+p+liwc(ling.)	.1953 (.1820,.2091)	.1882 (.1777,.1992)
MEAD c+p+liwc(psyc.)	.1913 (.1775,.2054)	.18550 (.1744,.1969)
MEAD c+p+liwc(pers.)	.1919 (.1783,.2051)	.1865 (.1756,.1972)

Results: Multi-Document

Results in DUC2002M

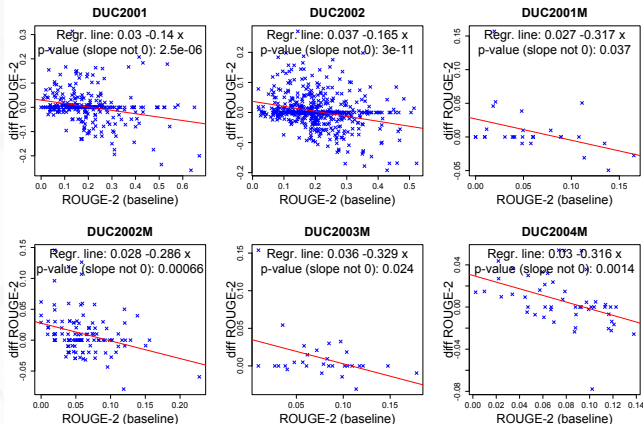
	ROUGE-2	ROUGE-SU4
default MEAD	.0684 (.0610,.0769)	.0950 (.0870,.1032)
random	.0355 (.0301,.0413)	.0710 (.0659,.0764)
lead-based	.0433 (.0369,.0504)	.0659 (.0601,.0716)
MEAD c+p tuned	.0610 (.0550,.0678)	.0963 (.0898,.1030)
MEAD c+p+liwc(all)	.0720 (.0643,.0810)	.1006 (.09371,.1083)
MEAD c+p+liwc(ling.)	.0711 (.0637,.0789)	.1047 (.0974,.1124)
MEAD c+p+liwc(psyc.)	.0626 (.0568,.0686)	.0931 (.0866,.0996)
MEAD c+p+liwc(pers.)	.0665 (.0594,.0736)	.0991 (.0911,.1069)

The **MEAD c+p+liwc(ling.)** summariser is consistently better than the baseline summarisers, however it does not achieve stat. sig. improvements, i.e., we think that it is working well for some cases but is **degrading the performance** for certain individual summarisation cases

For each summary we took ROUGE-2 score of a baseline summariser (MEAD c+p tuned) as an estimator of the difficulty to summarise the document or cluster

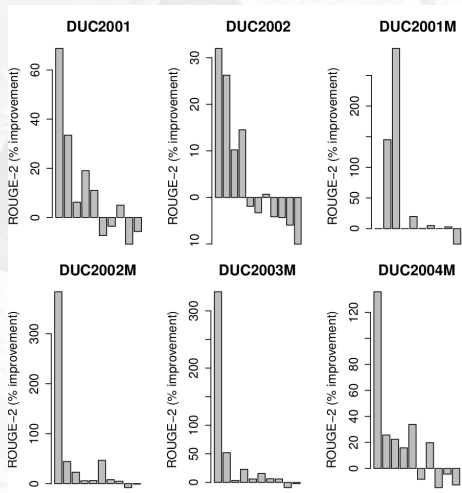
We computed the difference between the ROUGE-2 score of the **MEAD c+p+liwc(ling.)** summariser and the ROUGE-2 score of the baseline summariser.

Analysis: Results (i)



Our method has a tendency to work well for difficult summarisation cases (low ROUGE-2) and to be harming for easier summarisation cases.

Analysis: Results (ii)



Percentage of improvement of the MEAD c+p+liwc(ling.) with summaries binned on the baseline performance. performance

Having the weight of the different features we can conclude on the importance of each one in the summarisation process. In general, the summariser gives preferences to sentences that

- have quantifiers, prepositions, conjunctions, impersonal pronouns,
- lack personal pronouns, 1st person plural, and adverbs.

This fits well with some findings in the area of Psychology, related with the use of the language by people writing about real experiences

Our analysis suggests that driving the summarisers with LIWC features has implicitly fomented analytical extracts and extracts about real experiences.

CONCLUSIONS AND FUTURE WORK

Conclusions

- We have provided preliminary empirical evidence on the effect of psycholinguistic features in Automatic Text Summarisation.
- We defined a novel set of features –related to psychological dimensions– and injected them into a state-of-the-art summarisation system.
- We found that the summariser that includes linguistic LIWC dimensions is the best performing summariser. There are interesting connections between the occurrence of certain linguistic dimensions and types of writing and thinking.
- Our novel summarisation approaches are better suited for hard summarisation cases.

- We believe that there is room for further enhancement. For example, by applying feature selection to individually extract LIWC features from every subset of LIWC dimensions.
- We hope that our results serve as a basis to foster the discussion on how linguistic and psychological dimensions relate to sentence salience.
- Selective feature injection for summarisation: estimate the difficulty of summarising a given document or cluster and then decide whether or not to add the advanced features.

THANK YOU!

 @JPARAPAR

<http://www.dc.fi.udc.es/~parapar>