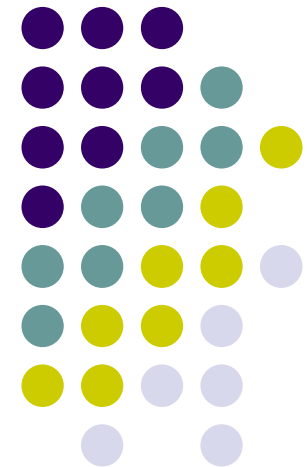


Opinion Mining and Agreement Identification in Forums

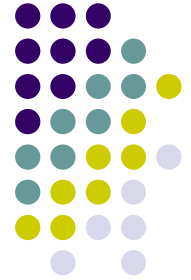
Anna STAVRIANOU,
Jean-Hugues CHAUCHAT

Université de Lyon - Laboratoire ERIC



FODOP, 27 Mai 2008

Plan



- Intérêt de la *fouille des données d'opinion* (Opinion Mining)
- Méthodes existantes
- Analyse des accords/désaccords

Intérêt de la fouille des données d'opinion



- Fouille de forum et de débat
- Comment l'opinion évolue dans la discussion
- Intensité des opinions identifiées
- Ironie et sarcasme
- Identification des accords/désaccords

Méthodes Existantes 1



Identification de l'orientation :

- Manuelle (effort important + coût)
- Basée sur corpus (dépendant du domaine)
- Basée sur dictionnaire (manque d'info contextuelle)

Méthodes Existantes 2



Approches	Seed list	Score	Info
Hatzivas. 1997			Conjoined adjs
Pang 2002			Absence/pres.
Wiebe 2000	√		Subj/Obj Phrase
Hu 2004	√		Features + OW
Ding 2007	√	√	
Esuli 2005	√	√	SentiWordNet
Kamps 2004	√	√	Path in WN
Ku 2007	√	√	+ topic detection
Turney 2003	√	√	PMI, LSA

Analyse des accords/désaccords



- L'identification de la "connotation" est différente de l'identification d'accord.

Exemple : " I agree that he is wrong " (accord, connotation: negative)

- *Objectif* :
 - Distinguer entre accord et désaccord (Forums, débats)
 - Extraire les arguments
 - Relations opinions - arguments
 - non opinions - caractéristiques

Approche

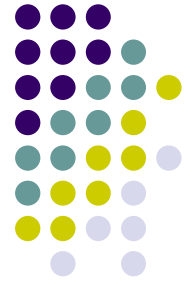


Est-ce que les méthodes actuelles (proposées pour trouver les relations entre opinions et caractéristiques) fonctionnent pour identifier les accords ?

Notre adaptation de ces méthodes :

1. Représentation du texte comme un ensemble de mots
2. Construction d'une seed list d'accord/désaccord
3. Calcul de la distance entre l'ensemble des mots du texte et la seed list

Représentation du texte



- Tokenisation (séparation des mots)
- PoS tagging (identification des noms, verbes ...)
- Extraction des verbes et noms
- Supression des Stopwords
- Repérage des Négations (I don't agree → I disagree)

Seed list d'accord/désaccord



- Catégories du *General Inquirer*
- Synonymes/Antonymes de *WordNet*

Distance entre les mots du texte et la seed list



- (Lin 1998) mesure
- Comparaison entre les 2 ensembles de mots
- Si la majorité des mots est plus proche de l'accord, alors le commentaire est annoté comme accord sinon désaccord

Petit Test



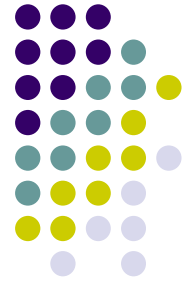
- www.englishforums.com
- 30 commentaires avec 77 mots en moyenne

Résultats



- L'accord est plus facilement identifié que le désaccord
- L'accord est souvent implicite et les mots utilisés ne peuvent pas être capturés par une seed list
- Quelques mots/expressions usuel(le)s ne sont pas toujours dans WordNet
- Les commentaires dans les forums n'expriment pas toujours l'accord ou le désaccord

Conclusion



- Fouille des données d'opinion est un champ qui présente beaucoup d'intérêt
- Techniques actuelles doivent être enrichies quand elles sont appliqués pour l'identification d'accord

Bibliographie



- Ding X., Liu B., “The utility of linguistic rules in Opinion Mining”, *SIGIR*, 2007
- Esuli A., Sebastani F., “SENTIWORDNET: A publicly available lexical resource for Opinion Mining”, *LREC*, 2005
- Hatzivassiloglou V., Mckeown, K.R., “Predicting the semantic orientation of adjectives”, *35th ACL*, 1997
- Hu M., Liu B., “Mining and summarizing customer reviews”, *KDD*, 2004.
- Kamps J., Marx M., Mokken R.J., De Rijke M., “Using WordNet to measure semantic orientations of adjectives”, *4th LREC*, 2004.
- Ku Lun-Wei, Chen Hsin-His, “Mining opinions from the web: beyond relevance retrieval”, *Journal of the American Society for Information Science and Technology*, 58(12), 2007
- Lin D., “Automatic retrieval and clustering of similar words”, *COLING-ACL*, 1998
- Pang B., Lee L., Vaithyanathan S., “Thumbs up? Sentiment classification using machine learning techniques”, *EMNLP*, 2002
- Stavrianou A., Andritsos P., Nicoloyannis N., “Overview and semantic issues of Text Mining”, *SIGMOD Record*, 36(3), 2007
- Turney P.D., Littman M.L., “Measuring praise and criticism: inference of semantic orientation from association”, *ACM TOIS* 21(4), 2003
- Wiebe J.M., “Learning subjective adjectives from corpora”, *AAAI*, 2000