



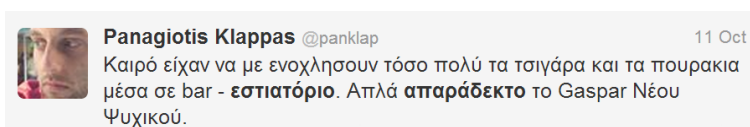
Διπλωματική Εργασία 2013-2014

Ανάλυση συναισθήματος σε δεδομένα κοινωνικών δικτύων με χρήση κατανεμημένων τεχνικών μηχανικής εκμάθησης

Εισαγωγή

Η εξέλιξη των κινητών συσκευών και η εξάπλωση των ασύρματων δικτύων έχουν συμβάλει δραστικά στην ευρεία χρήση των κοινωνικών δικτύων και υπηρεσιών. Όλο και περισσότεροι χρήστες χρησιμοποιούν τα κοινωνικά δίκτυα για να επικοινωνήσουν με τους φίλους τους, να δηλώσουν την παρουσία τους σε ένα μέρος, να εκφράσουν ιδέες και γνώμες, να μοιραστούν τη διάθεση ή τα συναισθήματά τους.

Η αξιοποίηση αυτού του θησαυρού πληροφοριών και η εξαγωγή γνώμης και συναισθήματος μέσα από αυτόν έχει μεγάλη σημασία τόσο για τους παροχείς προϊόντων και υπηρεσιών όσο και για τους καταναλωτές τους. Για παράδειγμα, αν μια τοπική επιχείρηση γνωρίζει την άποψη των καταναλωτών της σχετικά με τις παρεχόμενες υπηρεσίες της μπορεί να αποκτήσει πλεονέκτημα έναντι του ανταγωνισμού. Αντίστοιχα, ο καταναλωτής μπορεί να αποφύγει ένα μέρος που έχει λάβει κακές κριτικές στα κοινωνικά μέσα.



Σχήμα 1: Παράδειγμα tweet με αρνητικό συναίσθημα

Επιπλέον, με τον αριθμό των χρηστών κοινωνικών δικτύων να αυξάνεται ραγδαία (πάνω από 1 δις για Facebook), η αποτελεσματική διαχείριση και ανάλυση του τεράστιου όγκου των παραγόμενων δεδομένων απαιτεί κατανεμημένες λύσεις και αποτελεί μεγάλη πρόκληση για τον ερευνητικό κόσμο.

Στη διπλωματική αυτή θα ασχοληθούμε με την επεξεργασία δεδομένων σε μορφή κειμένου, τα οποία προέρχονται από κοινωνικά δίκτυα, με σκοπό την αυτόματη εξαγωγή του συναισθήματος που κρύβουν (sentiment analysis). Λόγω του μεγάλου όγκου των δεδομένων, για την αποτελεσματική διαχείριση κι επεξεργασία τους θα ακολουθηθούν κατανεμημένες προσεγγίσεις.

Σκοπός

Σκοπός της διπλωματικής είναι η ανάλυση του συναισθήματος σε μεγάλο όγκο κειμένου, προερχόμενου από κοινωνικά δίκτυα. Στο πλαίσιο αυτό θα μελετηθούν αρχικά οι πιο γνωστοί στη σχετική βιβλιογραφία αλγόριθμοι sentiment analysis οι οποίοι στηρίζονται σε τεχνικές μηχανικής εκμάθησης (machine learning) [3,5]. Στη συνέχεια θα επιλεχθούν οι αλγόριθμοι που ταιριάζουν καλύτερα στα ιδιαίτερα χαρακτηριστικά των κοινωνικών δικτύων (π.χ. [4]) και θα αναπτυχθούν σε καταναμημένο περιβάλλον. Αυτό θα γίνει με τη χρήση καταναμημένων εργαλείων όπως το Hadoop [1], το οποίο αποτελεί το πλέον διαδεδομένο framework για καταναμημένη επεξεργασία και το Apache Mahout [2], το οποίο προσφέρει μια βιβλιοθήκη για αλγορίθμους μηχανικής εκμάθησης πάνω από Hadoop. Τέλος, θα γίνει πειραματική αξιολόγηση των καταναμημένων αλγορίθμων ως προς την επίδοση, την ακρίβεια (precision) και την ανάκλησή τους (recall) με χρήση πραγματικών δεδομένων.

Επικοινωνία:

Νεκτάριος Κοζύρης, Καθηγητής nkoziris@cslab.ece.ntua.gr

Γιάννης Κωνσταντίνου, Μεταδ. Ερευνητής ikons@cslab.ece.ntua.gr

Κατερίνα Δόκα, Μεταδ. Ερευνήτρια katerina@cslab.ece.ntua.gr

Βιβλιογραφία

- [1] Apache Hadoop. <http://hadoop.apache.org/>. 2
- [2] Apache Mahout: Scalable machine learning and data mining. <http://mahout.apache.org/>. 2
- [3] Erik Boiy and Marie-Francine Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Information retrieval*, 12(5):526–558, 2009. 2
- [4] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, 2010. 2
- [5] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008. 2